



**João Vasco Espinhal
Otero da Costa**

**Monitorização e Geovisualização de Pesquisas *Web*
no Portal Sapo**



**João Vasco Espinhal
Otero da Costa**

**Monitorização e Geovisualização de Pesquisas *Web*
no Portal Sapo**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Comunicação Multimédia, realizada sob a orientação científica da Doutora Ana Veloso, Professora Auxiliar do Departamento de Comunicação e Arte da Universidade de Aveiro e sob a co-orientação do Doutor Rui Raposo, Professor Auxiliar Convidado do Departamento de Comunicação e Arte da Universidade de Aveiro.

dedicatória

Dedico este trabalho aos meus pais e irmão, por me suportarem e apoiarem durante os anos do meu percurso académico.

o júri

presidente

Doutor Óscar Emanuel Chaves Mealha

professor associado do Departamento de Comunicação e Arte da Universidade de Aveiro

Celso Manuel Batista Martinho

director de Tecnologia do Portal Sapo e Administrador não executivo da PT Inovação

Doutora Ana Isabel Barreto Furtado Franco de Albuquerque Veloso

professora auxiliar do Departamento de Comunicação e Arte da Universidade de Aveiro

Doutor Rui Manuel de Assunção Raposo

professor auxiliar convidado do Departamento de Comunicação e Arte da Universidade de Aveiro

agradecimentos

Quero agradecer a um conjunto de pessoas que ajudaram, directa ou indirectamente, na elaboração deste trabalho.

À Prof. Ana Veloso, ao Prof. Óscar Mealha e ao Prof. Rui Raposo, pelo apoio imprescindível prestado, não só ao longo deste ano, mas no decorrer do Mestrado.

Ao Prof. Carlos Santos, ao Prof. Doutor Luís Pedro, e ao Prof. Pedro Almeida, pelas ajudas prestadas e dúvidas tiradas.

À equipa Sapo, principalmente aos participantes do estudo, pela sua disponibilidade. Especial agradecimento ao Benjamin Júnior, Celso Martinho, Rui Gomes, Cláudio Valente e João Gonçalves, pelo acompanhamento e esclarecimento de algumas questões.

Ao Manuel Lima, pela grande ajuda que prestou, tanto na palestra leccionada na *Shift*, como pelos *e-mails* trocados respeitando tópicos da área da Investigação.

À minha namorada, pelo carinho e apoio, e paciência para me suportar.

Aos meus amigos, pela ajuda preciosa e convívios revigorantes.

Aos meus pais e ao meu irmão por tudo.

A todos eles o meu profundo agradecimento.

palavras-chave

Visualização de Informação, Geovisualização, Georreferenciação, pesquisas *Web*, Portal *Sapo*, *Sapo Broker*.

resumo

O presente trabalho pretende contribuir para a investigação realizada na área da Visualização de Informação. O projecto que este documento descreve consiste no desenvolvimento de um protótipo que consiga atribuir significado a dados recolhidos pelo sistema *Sapo Broker*. O *Sapo Broker* consiste numa aplicação de monitorização em tempo real, que recolhe dados relacionados com o Portal *Web Sapo*. Da gama ampla de dados recolhidos pelo sistema, pretende-se trabalhar com os que incidem especificamente sobre as pesquisas *Web* georreferenciadas realizadas no portal.

Na procura de resultados para a investigação encetada, apontam-se dois desafios principais: i) a tradução correcta dos dados em informação analisável segundo metáforas visuais apropriadas para o público-alvo, tecnologia e contexto de uso; ii) a definição das ferramentas de interacção necessárias para que a análise dos dados seja intuitiva e relevante.

keywords

Information Visualization, Geovisualization, Georeferencing, *Web* search, *Sapo Portal*, *Sapo Broker*.

abstract

This research aims to contribute to the existing research associated to the area of Information Visualization. The project this dissertation describes consists in the development of a prototype application capable of discerning and communicating meaning from data retrieved from the *Sapo Broker* system. The *Sapo Broker* system consists of a real-time monitoring application that retrieves user-related data from the Web Portal *Sapo*. From the extensive quantity of data collected, the following work concentrates primarily on data regarding *Web* search keywords and the Georeference from where those searches were performed. In order to conduct the following research, two main challenges have been identified: a) the use of visual metaphors capable of correctly translating data into useful and measurable information bearing in mind the applications target-audience, technology, and context of use; b) selecting the accurate interaction tools that can help the Information analysis process become relevant and more intuitive.

ÍNDICE

1. INTRODUÇÃO	1
1.1. TIPOLOGIAS DA INVESTIGAÇÃO	1
1.2. PROBLEMA DE INVESTIGAÇÃO	1
1.3. FINALIDADES E OBJECTIVOS	2
1.4. ESTRUTURA DO DOCUMENTO	3
2. ENQUADRAMENTO TEÓRICO	5
2.1 INTRODUÇÃO	5
2.1.1 <i>Visualização Científica</i>	8
2.1.2 <i>Visualização Geográfica</i>	9
2.2 INVESTIGAÇÃO E COMERCIALIZAÇÃO	11
2.2.1 <i>Many-Eyes</i>	11
2.2.2 <i>Perceptual Edge</i>	12
2.2.3 <i>Flowing Data</i>	12
2.2.4 <i>Visual Complexity</i>	12
2.2.5 <i>Spotfire</i>	12
2.2.6 <i>Tableau</i>	13
2.2.7 <i>XPLANE</i>	13
2.2.8 <i>Gapminder</i>	14
2.3 PERCEPÇÃO E COGNIÇÃO	14
2.3.1 <i>Percepção pré-consciente</i>	15
2.4 VISUALIZAÇÃO DE DADOS TEXTUAIS	18
2.5 PADRÕES DE DESIGN	22
2.5.1 <i>Padrões de Visualização</i>	23
2.5.2 <i>Padrões de Comportamento</i>	31
2.5.3 <i>Padrões de Interação</i>	35
2.6 USABILIDADE	36
2.6.1 <i>Usabilidade Universal</i>	37
2.6.2 <i>Estética da Informação</i>	38
2.6.3 <i>Técnicas de Avaliação</i>	39
3. INVESTIGAÇÃO EMPÍRICA DO ESTUDO DE CASO	41
3.1 APRESENTAÇÃO DA METODOLOGIA DE INVESTIGAÇÃO	41
3.2 PLANO DE ACÇÃO	42
3.3 CARACTERIZAÇÃO DO PÚBLICO-ALVO	43
3.4 CONTEXTUALIZAÇÃO TÉCNICA	44
3.5 DESIGN FUNCIONAL	44
3.5.1 <i>Pesquisas</i>	45
3.5.2 <i>Mapa geográfico</i>	48
3.5.3 <i>Estudo de Interface</i>	51
3.6 DESIGN TÉCNICO	53
3.6.1 <i>Recolha dos dados Broker</i>	54

3.6.2	<i>Visualização</i>	57
3.7	MÉTODOS E TÉCNICAS PARA A RECOLHA DE DADOS	68
3.7.1	<i>Técnicas utilizadas</i>	68
3.7.2	<i>Construção dos Instrumentos</i>	69
4.	APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	71
4.1	ANÁLISE QUANTITATIVA	71
4.1.1	<i>Representação Visual</i>	72
4.1.2	<i>Ferramentas de Interação</i>	77
4.2	ANÁLISE QUALITATIVA	79
4.2.1	<i>Expectativas</i>	80
4.2.2	<i>Representação visual</i>	80
4.2.3	<i>Ferramentas de interação</i>	80
4.2.4	<i>Perspectivas futuras</i>	81
5.	CONSIDERAÇÕES FINAIS	83
5.1	ANÁLISE DA PERGUNTA DE INVESTIGAÇÃO	83
5.2	LIMITAÇÕES DO ESTUDO	85
5.2.1	<i>Limitações na implementação</i>	86
5.2.2	<i>Limitações no processo de avaliação</i>	87
5.3	ANÁLISE DO PROTÓTIPO DESENVOLVIDO	88
5.4	PERSPECTIVAS DE TRABALHO FUTURO	89
5.5	REFLEXÕES FINAIS	91
6.	BIBLIOGRAFIA	93
7.	ANEXOS	97

ÍNDICE DE FIGURAS

FIGURA 1 - EXEMPLO DA APLICAÇÃO <i>MICROSOFT EXCEL</i>	6
FIGURA 2 - O FAMOSO MAPA DE MINARD DAS TROPAS DE NAPOLEÃO (TUFT, 1990).	7
FIGURA 3 - EXEMPLO DE UMA VISUALIZAÇÃO CIENTÍFICA, A APLICAÇÃO <i>VISIBLE BODY</i>	8
FIGURA 4 - EXEMPLO DE UMA GEOVISUALIZAÇÃO: <i>BRITAIN FROM ABOVE</i>	10
FIGURA 5 - EXEMPLO DA VISUALIZAÇÃO DE TEXTO, ATRAVÉS DA APLICAÇÃO <i>WORDLE</i>	18
FIGURA 6 - <i>ELASTIC TAG MAPS</i> , DE MORITZ STEFANER.	20
FIGURA 7 - DE UM AGREGADOR DE NOTÍCIAS EM FORMATO DE MAPA DE ÁRVORE: <i>NEWSMAP</i>	21
FIGURA 8 - O EXEMPLO DE UM DOS ECRÃS DA APLICAÇÃO <i>WE FEEL FINE</i>	22
FIGURA 9 - EXEMPLO DE UM BUBBLECHART, UMA DAS FERRAMENTAS DE VISUALIZAÇÃO DO <i>GAPMINDER</i>	24
FIGURA 10 - <i>LISTENING HISTORY – LAST.FM</i> , UM EXEMPLO DE UMA APLICAÇÃO QUE USA A TIPOLOGIA DO <i>STREAM GRAPH</i>	25
FIGURA 11 - UM EXEMPLO DE UM <i>STARPLOT</i>	26
FIGURA 12 - A APLICAÇÃO <i>SHAPE OF A SONG</i> , DE MARTIN WATTENBERG 27	27
FIGURA 13 - EXEMPLO DE UM DIAGRAMA EM ESTRELA (<i>STARTREE</i>).....	28
FIGURA 14 - EXEMPLO DE UM CÍRCULO DE RELAÇÕES – PROJECTO <i>CLUSTERBALL: VISUALIZING WIKIPEDIA</i>	29
FIGURA 15 - CRESCIMENTO DAS LOJAS <i>WALLMART</i> NOS EUA AO LONGO DOS ANOS.....	30
FIGURA 16 - APLICAÇÃO <i>GENREBALLUNG</i> , DE CHRISTOPHER ADJEI E NILS HOLLAND-CUNZ.....	31
FIGURA 17 - UMA APLICAÇÃO DA EMPRESA <i>SPOTFIRE</i>	33
FIGURA 18 – A APLICAÇÃO <i>GOOGLE MAPS</i>	35
FIGURA 19 – APLICAÇÃO <i>MAP OF THE MARKET</i> , DO PORTAL ONLINE <i>SMARTMONEY</i> 37	37
FIGURA 20 – EXEMPLO DA TIPOLOGIA <i>SUNBURST</i> , DO PROGRAMA <i>DOCUBURST</i> 39	39
FIGURA 21 - PRIMEIRA EXPERIÊNCIA NO PROCESSING: LISTAGEM DE PALAVRAS COM NÚMEROS DE OCORRÊNCIAS.....	45
FIGURA 22 – PROJECTO <i>43THINGS</i>	46
FIGURA 23 – RESPOSTA IMPLEMENTADA BASEADA NO PROJECTO <i>43THINGS</i>	47
FIGURA 24 – ESQUEMA VISUAL ADOPTADO PARA A VISUALIZAÇÃO.	48
FIGURA 25 – UMA DAS PRIMEIRAS EXPERIÊNCIAS COM AS DISTRIBUIÇÕES GEOGRÁFICAS.	49
FIGURA 26 – UMA DAS TIPOLOGIAS A SEREM ESTUDADAS, QUE PERMITEM VER MAIS DETALHES 50	50
FIGURA 27 – CONTEÚDO FINAL DISPONIBILIZADO NOS DETALHES À MEDIDA DE CADA OVO.....	50
FIGURA 28 – CÍRCULO <i>POP-UP</i> , RELATIVO AO OVO QUE FOI CLICADO.	51
FIGURA 29 – PRIMEIRAS FERRAMENTAS DE INTERACÇÃO A SEREM CONCEPTUALIZADAS.	52
FIGURA 30 – SOLUÇÃO ESCOLHIDA PARA AS FERRAMENTAS PARA SELECIONAR O TOP, OU O INTERVALO DE SELECÇÃO 52	52
FIGURA 31 – AS DUAS FERRAMENTAS DE INTERACÇÃO: A ESCALA TEMPORAL E A CAIXA DE TEXTO.	53
FIGURA 32 – APARÊNCIA FINAL DA CONCEPTUALIZAÇÃO, ADOPTADA PARA A IMPLEMENTAÇÃO DO PROTÓTIPO.	53
FIGURA 33 – DIFERENTES ESTADOS PARA AS VARIÁVEIS X0 59	59
FIGURA 34 – EQUAÇÕES DE MOVIMENTO (HORIZONTAL E VERTICAL) DOS OVOS 60	60
FIGURA 35 – SÉRIE GERADORA DA ESCALA PERCENTUAL DO NÚMERO DE PESQUISAS GEORREFERENCIADAS.	62
FIGURA 36 – PROCESSO DE EXECUÇÃO DO CÍRCULO DE POP-UP DE UM OVO.....	64
FIGURA 37 – TOP PESQUISAS.	65
FIGURA 38 – INTERVALO DE PESQUISAS.	65
FIGURA 39 – ESCALA TEMPORAL.	66
FIGURA 40 – SEGMENTOS TEMPORAIS.	66
FIGURA 41 – FILTRAGEM DA PALAVRA-PESQUISA 67	67
FIGURA 42 - DIFERENCIAÇÃO DAS DIFERENTES FERRAMENTAS DE INTERACÇÃO DA APLICAÇÃO.	77

ÍNDICE DE GRÁFICOS

GRÁFICO 1 – RESOLUÇÃO DA APLICAÇÃO.	72
GRÁFICO 2 – METÁDORA VISUAL - ESTÉTICA.	73
GRÁFICO 3 – METÁFORA VISUAL – REPRESENTATIVIDADE.	73
GRÁFICO 4 – DENSIDADE DA VISUALIZAÇÃO.	74
GRÁFICO 5 – TAMANHO DOS OVOS DA VISUALIZAÇÃO.	74
GRÁFICO 6 – MODO DE VER MAIS DETALHES.	75
GRÁFICO 7 – DIFICULDADE EM ENCONTRAR ELEMENTOS DE INTERACÇÃO.	76
GRÁFICO 8 – DIFICULDADE EM ENCONTRAR DETERMINADAS PALAVRAS-PESQUISAS.	76
GRÁFICO 9 – FEEDBACK RECEBIDO QUANDO SE ALTERAM ESTADOS DA VISUALIZAÇÃO.	78
GRÁFICO 10 – GRAU DE IMPORTÂNCIA DAS FERRAMENTAS DE INTERACÇÃO.	79

ÍNDICE DE TABELAS

TABELA 1 – REPRESENTAÇÃO ESQUEMÁTICA DO CONJUNTO DOS ARRAYS RELATIVOS AOS DADOS DE PESQUISA	56
TABELA 2 – FORMATO DO FICHEIRO COM O TOP DE PESQUISAS.	57

1. INTRODUÇÃO

1.1. TIPOLOGIAS DA INVESTIGAÇÃO

A presente investigação caracteriza-se como sendo um estudo de caso. Envolve sobretudo uma componente prática, que se traduz no desenvolvimento de um protótipo funcional, aplicado ao contexto de uso específico do público-alvo escolhido.

Implica uma pesquisa bibliográfica exaustiva sobre conceitos teóricos relevantes, e um estudo de campo onde se pretende aferir as necessidades do utilizador – antes e depois da concepção do protótipo. O primeiro momento do estudo de campo tem o objectivo de contribuir para a formulação dos objectivos da investigação, e o segundo é orientado para a avaliação do protótipo.

No que diz respeito à obtenção e tratamento de dados, este estudo de caso é principalmente qualitativo, incidindo numa análise de conteúdo. Também alguns dados quantitativos foram recolhidos acerca de tópicos considerados relevantes, com o objectivo de complementar a análise qualitativa da investigação.

1.2. PROBLEMA DE INVESTIGAÇÃO

O *Sapo Broker* é um sistema do Portal Sapo, que monitoriza um grande número de variáveis, em tempo real, sobre os eventos que acontecem dentro deste: o número de cliques efectuados em diferentes zonas do portal, quais os conteúdos mais procurados, de que distrito do país é feito este acesso, entre outros.

Estes dados têm um grande valor, no sentido em que podem ser indicadores das preferências dos utilizadores, num mercado que está em constante mudança; ao longo do tempo, dão-se várias mudanças de comportamento, que devem ser acompanhadas pelos dinamizadores do portal, procurando-se adaptar ao que está em voga.

No entanto, muita da informação disponibilizada por este sistema não é ainda utilizada na sua totalidade. Existindo já uma aplicação que utilize esta informação, esta ainda não consegue aproveitar todo o seu potencial; falta um sistema que armazene estes dados para a realização de análises detalhadas, e que ao mesmo tempo possibilitem que estas análises tenham uma forma menos tecnicista e mais intuitiva.

Quando esta análise é difícil de empreender, requer usualmente especialistas que não conseguem depois transmitir a informação para o resto do *staff*, num formato que seja cativante e de fácil leitura; tal resulta muitas vezes em que essa informação seja lida superficialmente ou posta de parte (Rosling, Ronnlund, & Rosling, 2004).

Esta tendência poderá ser contrariada através da utilização de sistemas de análise de dados, que sejam sobretudo sistemas visuais. Fazendo uso de metáforas espaciais, familiares ao olho humano, estes sistemas são capazes de conferir corpo a dados abstractos, enquanto utilizam ferramentas de interacção que permitem a sua manipulação directa.

Tendo isto em conta, esta investigação pretende aproveitar alguns dos dados recolhidos pelo *Sapo Broker*, especificamente os dados sobre as pesquisas efectuadas no portal, associadas à região do país de onde são efectuadas. Com base nestes dados, e no contexto da instituição, procurar-se-ão chegar a novas formas de visualizar estes dados, de maneira a transmitir mais intuitivamente nova informação. Da mesma forma, procurar-se-ão averiguar quais as análises dos dados mais interessantes para serem efectuadas através destes dados, e quais as ferramentas interacção mais indicadas para as operacionalizarem.

Pretende-se assim desenvolver pontos que estão em falta no actual sistema de representação dos dados de pesquisas, adequando-se mais ao público-alvo a que se destina: o *staff* da redacção do portal Sapo. Com base nestes pressupostos surge então a seguinte pergunta de investigação:

Que proposta de metáfora visual poderá ser conceptualizada e implementada com base nos dados do broker resultantes das pesquisas realizadas no portal Sapo?

Tendo em conta a pergunta de investigação formulada, podemos estipular algumas hipóteses sobre maneiras possíveis de abordar esta questão:

- Através da representação de uma amostra de dados, que permita de um modo rápido verificar picos de pesquisa em tempo real, dando pouco ênfase às suas características particulares. Cada dado de pesquisa poderá ser representado por um corpo individual, cuja diferenciação dos outros corpos poderá dar-se a um nível morfológico e/ou espacial.
- A distribuição geográfica das pesquisas efectuadas poderá ser representada numa janela temporal, que transmita imediatamente quais os distritos do país com uma maior afluência de pesquisas, sendo possível ao utilizador explorar os dados, permitindo-lhe ver um nível personalizado de detalhes.

1.3. FINALIDADES E OBJECTIVOS

Como finalidade para esta investigação, pretende-se contribuir para o desenvolvimento dos conhecimentos académicos existentes na área da Visualização de Informação, procurando contribuir para a formulação de novos paradigmas de análise de dados, que inspirem novas formas de pensar e interagir visualmente.

Do ponto de vista funcional, pretende-se criar uma ferramenta de prospecção suportada numa análise visual de dados, que possa facilitar a descoberta de novas tendências ou destaques noticiosos através dos registos de pesquisa em motores *on-line*.

Perspectiva-se ainda que posteriormente seja possível a generalização desta abordagem para outras áreas de actividade, sectores, e instituições (que possam envolver áreas como o *e-marketing*, *e-commerce* ou *e-learning*).

Tendo em conta a problemática e finalidade definidas para esta investigação, é proposto o seguinte objectivo geral:

- Apresentar metáforas de análise visuais e interactivas, que representem a distribuição das pesquisas no portal Sapo e a sua respectiva distribuição por distrito;

De seguida, em seguimento do objectivo geral, são propostos os seguintes objectivos específicos.

- Permitir a manipulação directa dos parâmetros de análise:
 - Palavras-chave a procurar;
 - Período de tempo para efectuar a leitura e análise dos dados;
 - Equalização da distribuição de ocorrências;
 - Número de resultados;
- Monitorizar e alertar permanentemente para a ocorrência de desvios de padrão significativos num motor de pesquisa por intervalo de tempo.

1.4. ESTRUTURA DO DOCUMENTO

O presente documento encontra-se dividido em 7 capítulos distintos: Introdução, Enquadramento Teórico, Investigação Empírica do Estudo de Caso, Apresentação, Análise e discussão dos Resultados, Comentários Finais, Bibliografia e Anexos.

1. A introdução pretende introduzir em que âmbito se enquadra a presente investigação. Para tal apresenta a problemática da investigação, esclarecendo o seu contexto de uso. Também define as finalidades e objectivos, fios condutores da investigação, assim como descreve quais as tipologias em que esta se insere. Por fim enumera as várias secções que estruturam este documento (a presente secção).
2. O Enquadramento Teórico reúne uma contextualização teórica, através da descrição da área de investigação, com a explicação de alguns conceitos relacionados que têm maior importância para a concretização dos objectivos da investigação. Neste capítulo são apresentadas definições de conceitos relevantes, projectos desenvolvidos, mecanismos relacionados, padrões de implementação mais comuns, e questões de usabilidade, dentro da área de estudo da investigação.
3. O capítulo que se intitula Investigação Empírica do Estudo de Caso tem o objectivo de descrever o contexto de uso da investigação, e para que tipo de resultados se dirige. Começa-se por apresentar qual a metodologia que vai ser adoptada no decorrer da investigação.

Segue-se a enumeração das fases sobre as quais esta se divide, e de que maneira é que estas estão interligadas. Também a caracterização do público-alvo, assim como da amostra seleccionada, estão retratados neste capítulo. De seguida é descrita a fase de conceptualização e implementação, que descreve todo o processo de desenvolvimento do protótipo da investigação. Por fim, são descritos os métodos e técnicas usadas para a recolha de dados, necessárias para a avaliação do protótipo junto do público-alvo.

4. Segue-se o capítulo da Apresentação, análise e discussão dos Resultados, que se separa por análise quantitativa e qualitativa, que engloba um conjunto de gráficos representativos das respostas obtidas e os seus respectivos comentários descritivos.
5. Os comentários finais englobam a reflexão crítica sobre os resultados obtidos, assim como de todo o processo da investigação. Insere-se neste capítulo a análise das questões de investigação, a apresentação das limitações encontradas no decorrer do estudo, a análise do protótipo desenvolvido, as perspectivas sobre o trabalho futuro que possa ser efectuado com base nesta investigação, e por fim as conclusões finais.
6. No seguinte capítulo, da Bibliografia, são listadas as referências bibliográficas usadas na investigação.
7. Por fim, é disponibilizada uma secção dos Anexos, que engloba anexos impressos e a referência e outros anexos inseridos no CD digital, que é disponibilizado juntamente com o presente documento.

2. ENQUADRAMENTO TEÓRICO

Numa fase inicial da investigação, foi realizada uma pesquisa bibliográfica exaustiva sobre a área de investigação, cujos resultados mais relevantes são aqui apresentados.

2.1 INTRODUÇÃO

Com o rápido crescimento da capacidade de memória e processamento dos sistemas computadorizados, é possível armazenar grandes quantidades de dados, dos tipos mais variados, e realizar operações de análise estatística mais complexas.

Porém, ao mesmo tempo que se deu este crescimento técnico, não foi dada especial importância a como tornar toda esta informação compreensível e interessante para o público geral, com o objectivo de facilitar a descoberta de novos significados, de forma clara, precisa, e útil (Few, 2007).

Números em tabelas são insignificantes para a maioria das pessoas. Contudo, o interesse do público multiplicar-se-ia se toda esta informação recolhida fosse, de algum modo, transmitida livremente pela internet através de interfaces gráficas interactivas e divertidas (Rosling, Ronnlund, & Rosling, 2004). Infelizmente, a maioria das pessoas encara a análise de dados como uma tarefa enfadonha, efectuada sobretudo através de gráficos dos tipos mostrados por programas tais como o *Microsoft Excel* (Figura 1) (Danziger, 2008).

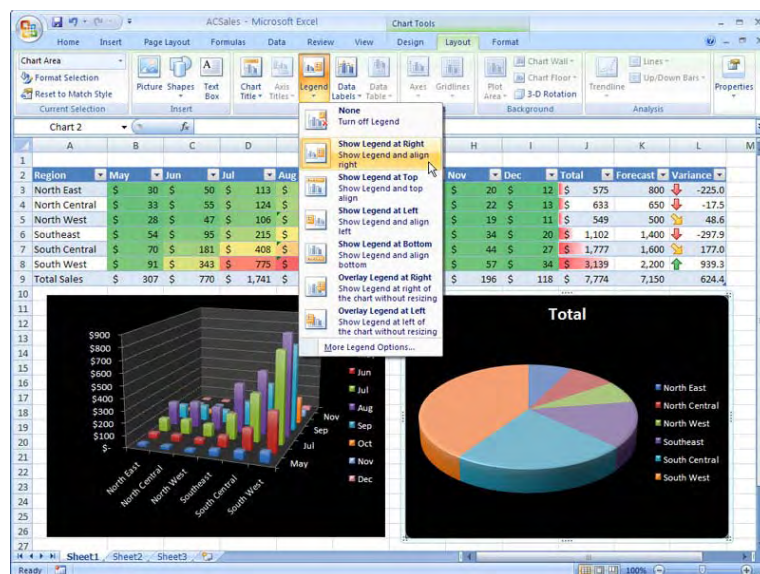


FIGURA 1 - EXEMPLO DA APLICAÇÃO MICROSOFT EXCEL¹. PARECE DAR A ENTENDER QUE A IDEIA DE PROGRESSO NA VISUALIZAÇÃO DE INFORMAÇÃO RESUME-SE APENAS A EMBELEZAR O QUE JÁ EXISTE (FEW, 2007).

Várias áreas sobre Visualização começaram a surgir, procurando lidar com estes sistemas informáticos de análise de dados, tendo todas elas a mesma herança comum - o Design de Informação (*Information Design*), mais especificamente o ramo da Informação Gráfica (*Infographics*), que procura desde há algum tempo novas formas de transformar e organizar a informação de um modo visual.

Dentro desta área foram realizados ao longo do tempo projectos que ainda hoje são marcantes sendo anteriores ao século XX - foram revelando o potencial da informação visual em lidar com dados quantitativos. Alguns deles são apresentados de seguida:

- *William Playfair* (1786) procurou formas de tornar dados económicos mais fáceis de compreender para políticos através dos primeiros gráficos de barras (Tuftes, 1983, p. 40).
- Para combater o surto de cólera que se abateu num distrito de Londres (1854), o médico *John Snow* criou um mapa que relacionava geograficamente as zonas com maior número de casualidades com as zonas públicas de distribuição de água. Permitiu assim determinar o padrão de propagação da doença, considerada antes como sendo transmissível pelo ar (Spence, 2007).
- *Florence Nightingale* (1858) escreveu um relatório ao governo britânico, onde demonstrava a melhoria das condições hospitalares, através da formulação de um Diagrama de Área Polar, que traduzia bem a redução dramática das mortes de pacientes ao longo dos meses. Este diagrama foi o precursor dos gráficos circulares, ou gráficos em anel (Spence, 2007).
- Uma das visualizações mais populares é o *Mapa da Marcha do exército de Napoleão à Rússia* (Figura 2), feita por *Charles Minard* (1861). Através deste impressionante trabalho, é possível discernir facilmente o caminho efectuado pelo exército, correlacionado com as perdas dos seus soldados (Danziger, 2008).

¹ Retirado de: <http://www.winplanet.com/img/screenshots/scr-office2007-4.jpg>

As datas dos dias de consulta dos *links* referenciados neste documento não puderam ser aferidas. No entanto, a validade de todos estes *links* foi comprovada pelo autor no dia 15 de Junho de 2009.

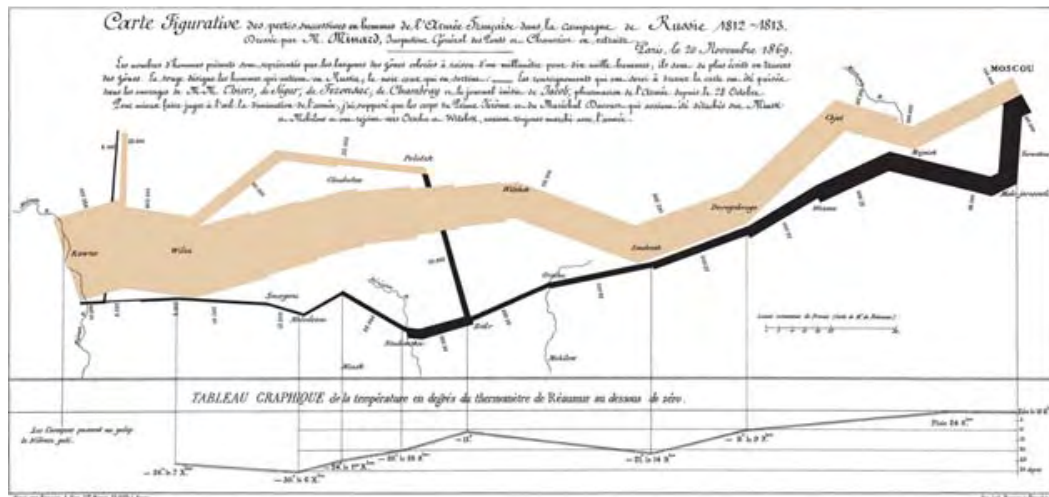


FIGURA 2 - O FAMOSO MAPA DE MINARD² DAS TROPAS DE NAPOLEÃO (TUFTE, 1990).

- Em 1931, Harry Beck, revolucionou os mapas, principalmente nos sistemas de transportes. Considerou que quando se anda de metro não interessa onde se está, pois não há paisagem a contemplar. Interessa sim de onde se parte e para onde se quer ir. Criou assim um mapa que podia ser facilmente lido, independentemente do número de linhas de metro que incorporasse (Spence, 2007).

Apesar de crucial no estudo da visualização, a Informação Gráfica não aborda questões relativas à interactividade dos sistemas de informática. Foi necessária a introdução de uma nova área temática que abordasse este tipo concreto de sistemas, mantendo objectivos semelhantes: a Visualização de Informação.

A Visualização de Informação consiste na representação e manipulação de dados abstractos, mediada por computador, com o objectivo de gerar e comunicar novo conhecimento.

O que distingue principalmente os conceitos de Informação Gráfica e Visualização de Informação é que enquanto a primeira representa visualmente informação de uma forma estática, e como tal mais desligada do factor tecnológico, a segunda caracteriza-se pelos seus sistemas serem interactivos, cujas representações visuais são dinâmicas, podendo lidar com conjuntos mais complexos de dados. Para além das questões da Visualização de dados, a Visualização de Informação tem de lidar com a componente da interacção Humano-Computador, e com as questões que esta possa implicar.

A definição original mais adoptada manteve-se mais ou menos constante na última década:

“Information visualization is the use of computer-supported, interactive, visual representations of abstract data in order to amplify cognition” (Card, Mackinlay & Shneiderman, 1999, p. 7).

Actualmente, também se reconhece a Comunicação do conhecimento gerado como uma nova componente dos sistemas de Visualização de Informação, sendo acrescido o factor social às Visualizações.

A área da Visualização de Informação procura sobretudo responder a perguntas do tipo: “Como transformar a análise de dados num processo intuitivo e interessante?”, “Como se visualizam dados destituídos de alguma forma física?”, e “Que tipo de interacção pode auxiliar a descoberta de novo conhecimento?”

Estes tipos de perguntas estão inter-ligados: uma aplicação intuitiva implica que a informação que nela é apresentada seja não só facilmente percebida mas também convenientemente manipulada, de maneira a auxiliar o utilizador na realização de tarefas da análise dos dados.

² Retirado de: <http://www.latebytes.nl/archives/2008/04/minard.html>

Os dados utilizados nestes sistemas entendem-se como dados abstractos, pois são aqueles que não têm associadas imagens representativas, tais como dados financeiros, conceitos abstractos, informação de negócios, por ex. Estes dados representam fenómenos que não são à partida fáceis de imaginar através do pensamento espacial a que estamos habituados. O objectivo da Visualização de Informação é otimizar o uso da nossa capacidade de raciocínio visual ao lidar com este tipo de informação (Chen, 2002).

2.1.1 VISUALIZAÇÃO CIENTÍFICA

O conceito de Visualização de Informação não deve ser confundido com o conceito de Visualização Científica: embora estejam as duas áreas interligadas dentro da Informação gráfica, lidam com diferentes tipos de informação e problemas, e têm aplicações diferentes.

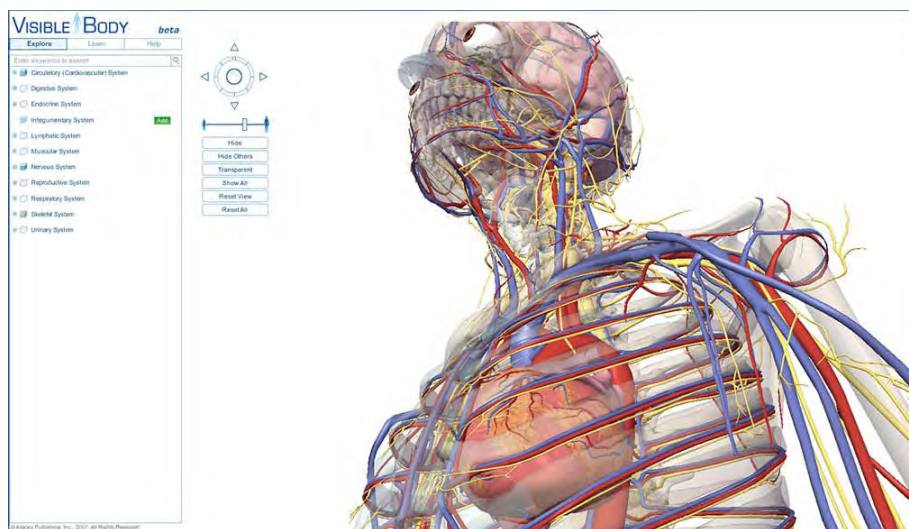


FIGURA 3 - EXEMPLO DE UMA VISUALIZAÇÃO CIENTÍFICA, A APLICAÇÃO *VISIBLE BODY*³, PRODUTO COMERCIAL DA EMPRESA ARGOSY PUBLISHING.

A Visualização Científica trabalha com dados que têm um carácter mais objectivo, relacionados com elementos científicos, directamente associadas a fenómenos da natureza. Está normalmente associada a representações tridimensionais de sistemas de biologia (Figura 3), física, e meteorologia, e é sustentada por teorias que conseguem quantificar previamente o significado de cada dado, não deixando espaço para qualquer erro de interpretação ou de representação.

Ou seja, a Informação Científica tem sempre algum tipo de modelo real, do qual se podem extrair os moldes necessários para a sua modelação geométrica, com as mesmas propriedades do fenómeno subjacente.

Por sua vez, a Visualização de Informação, lida com dados abstractos, sem terem uma forma física concreta. Não é guiada por directrizes e fórmulas que formatem as representações, pois a informação com que lida não tem atributos visuais. Cabe então aos *designers* de informação uma tarefa mais difícil, na medida em que são eles que têm de inventar modelos de visualização, sustentados por critérios subjectivos – que frequentemente variam conforme o indivíduo (Spence, 2007).

³ Retirado de: <http://www.visiblebody.com/start>

2.1.2 VISUALIZAÇÃO GEOGRÁFICA

A Geovisualização é outra área com particular destaque para esta investigação, ligada à Visualização de Informação, e utiliza também dados relativos a elementos físicos, nomeadamente referências geográficas.

À semelhança da Visualização Científica, a Geovisualização, ou Visualização Geográfica, guia-se segundo referências reais, em que as suas visualizações são sob alguma forma de mapa, representativo de uma zona geográfica específica, sempre dependente de um contexto espacial pré-existente. Estes mapas estão sempre ligados a um local da superfície terrestre, podendo esta informação ser recolhida através de várias formas, como imagens de satélite e bases de dados georreferenciados (Goodchild, 2006).

Os projectos de Geovisualização associam normalmente 3 tipos de variáveis: os lugares existentes na superfície da Terra, os seus atributos e as suas relações. Os lugares são conseguidos através de um sistema de coordenadas, tais como a latitude e a longitude, associados a pontos específicos pré-existent; os atributos dizem respeito a dados abstractos, que variam consoante as especificações dos projectos (como população, índice de poluição, etc.), que estão ligados a lugares geográficos específicos; por fim as relações comparam os diferentes atributos, através de vários critérios, tais como proximidade e conectividade.

Uma distinção importante a fazer é a que existe entre Geovisualização e a sua precursora, a Ciência da Informação Geográfica. A primeira focaliza-se na informação espacial e em como é que esta se relaciona com outro tipo de dados de natureza não-espacial, enquanto que a segunda incide na descrição detalhada das referências reais dos lugares geográficos correspondentes.

Os primeiros sistemas *GIS* (Sistemas de Informação Geográfica) começaram a surgir nos anos 60, sendo um dos primeiros, o *CGIS* (*Canada Geographic Information System*). Este sistema surgiu como forma de agregar a grande quantidade de mapas já existentes do país e, através deles, inquirir o índice de aproveitamento produtivo das suas regiões (Goodchild, 2006).

O percurso da Geovisualização está relacionado com a própria evolução da Internet. Esta relação começa concretamente com a monitorização do tráfego da rede *ARPANET*, do Departamento de Defesa norte-americano; através da construção dos primeiros mapas geográficos digitais para monitorizar a expansão desta rede pelo território do país ao longo dos anos.

Projectos deste tipo evoluíram até hoje, quando existe um enfoque especial no tráfego da *World Wide Web*, devido ao seu grande peso no tráfego global da Internet (Figura 4); tendo em conta áreas como o comércio electrónico e serviços de *ISP* (*Internet Service Providers*), surge a necessidade de compreender a distribuição geográfica dos seus acessos (Chen, 2006).



FIGURA 4 - EXEMPLO DE UMA GEOVISUALIZAÇÃO: *BRITAIN FROM ABOVE*⁴, QUE NESTA FIGURA RETRATA O TRÁFEGO AÉREO SOBRE O PAÍS DA INGLATERRA.

Com o surgimento da tendência generalizada da partilha social na *Web*, também nestes sistemas foram introduzidos mecanismos de criação e partilha da informação. Vários serviços públicos *online* permitem aos seus utilizadores criarem as suas Visualizações Geográficas através bases de dados por eles fornecidos, e conforme as suas especificações (tal como o sistema *Google Maps*⁵).

Existe quem defenda que a Geovisualização é uma das principais influências da Visualização de Informação (Chen, 2006), assim como existe quem afirma que a Geovisualização adopta conceitos de várias áreas, entre as quais a Visualização de Informação (MacEachren & Kraak, 2001):

“Geovisualization integrates approaches from scientific visualization, (exploratory) cartography, image analysis, information visualization, exploratory data analysis (EDA) and GIS to provide theory, methods and tools for the visual exploration, analysis, synthesis and presentation of geospatial data” (MacEachren & Kraak, 2001, p. 2).

Na verdade, é particularmente difícil dizer onde acaba uma e começa a outra, sendo mais verosímil considerar que as duas se intersectem em muitos pontos (Spence, 2007).

De um ponto de vista mais pragmático, podemos afirmar que a Geovisualização está inserida na Visualização de Informação, distinguindo-se desta por empregar nos seus sistemas referências a espaços geográficos terrestres.

Frequentemente, os sistemas Visualização de Informação utilizam também metáforas espaciais, seguindo os mesmos elementos gráficos dos mapas convencionais, mas sem a utilização necessária de referências geográficas reais: mapeiam-se espaços de informação, em vez de espaços geográficos. A eficiência destas metáforas espaciais está muitas vezes ligada à relação que existe entre os espaços geográficos e a experiência do utilizador (Slocum, et al., 2001). Este princípio é observado na aplicação *ThemeRiver™* (Havre, Hetzler & Nowell, 1999), onde a ideia de fluidez e dinâmica temporal é traduzida visualmente através dos “rios” de informação, que são reconhecidos pelo utilizador.

Seguindo os mesmos propósitos da Visualização de Informação, muitas investigações do âmbito da Geovisualização procuram explorar novos paradigmas de representação e manipulação da informação, com

⁴ Disponível em: <http://www.bbc.co.uk/britainfromabove/stories/visualisations/planes.shtml>

⁵ Disponível em: <http://maps.google.com/>

o objectivo de fomentarem novas descobertas. A introdução de novas metáforas visuais incute novas formas de pensar, contrariando a limitação criativa imposta pelos hábitos cognitivos (Kraak, 2002).

Um estudo interessante desta exploração criativa consiste na reutilização do já referido mapa de *Minard* (Figura 2, página 7) como base geradora de novos mapas alternativos, onde se adoptam novos pontos de vista sobre os mesmos dados; estas transformações contam-se entre: mudanças dos eixos de coordenadas e a apresentação das múltiplas representações visuais resultantes por cada mudança, a introdução de escalas de cor em elementos diferentes ou a transposição do mapa para uma plano a 3 dimensões. É possível verificar que através da mudança das perspectivas da visualização, assim como da adição de elementos interactivos, é possível chegar a novas descobertas, através de um mesmo conjunto de dados (Kraak, 2002).

No âmbito desta investigação, a componente da Geovisualização está presente nos seus objectivos, mas neste enquadramento teórico incidir-se-á mais de um modo geral na área da Visualização de Informação, onde já estarão integrados alguns pontos da geovisualização. Apesar de crucial, considera-se que a componente dos mapas não é a mais importante para a investigação, pois os dados que vão ser tratados no projecto não estão ligados a uma Georreferenciação muito detalhada (cujo grau de detalhe centra-se apenas nos distritos do país).

2.2 INVESTIGAÇÃO E COMERCIALIZAÇÃO

Esta secção tem como objecto enquadrar a temática desta investigação num contexto internacional, e discernir algumas das iniciativas mais relevantes ligadas à Visualização de Informação. O principal objectivo que as une é a divulgação e promoção desta área de investigação - sendo esta relativamente nova, é importante criar formas de divulgar novas questões ou inovações que lhe dizem respeito.

De uma forma ou de outra, estes projectos foram respondendo a novos desafios - enquanto alguns vão deixando o caminho aberto para futuros melhoramentos e o lançamento de novos problemas; outros centram-se na comercialização de sistemas de Visualização de Informação, ao encontrarem e desenvolverem um nicho de mercado em crescimento.

2.2.1 MANY-EYES

O portal *Many-Eyes*⁶ propõe-se a disponibilizar, sob o formato de comunidade virtual, um conjunto de ferramentas interactivas de Visualização, que possam ser utilizadas por um público alargado - possibilitando a cada pessoa uma forma mais pessoal de analisar dados.

Lançado em Janeiro de 2007, o site desenvolvido por *Martin Wattenberg*, *Fernanda Viegas* e outros (da *IBM Research*) permite o *upload* de dados e a sua Visualização digital. Além disso, permite a sua discussão aberta dentro da comunidade, através de comentários efectuados pelos utilizadores. Disponibiliza um largo conjunto de ferramentas avançadas de Visualização, tais como Mapas de árvore, Diagramas de Redes e Mapas.

Também é possível a utilização de bases de dados de outros utilizadores - assim como a sua modificação - potencializando assim o trabalho comunitário. Existe a preocupação especial em tornar o site o menos complexo possível para pessoas que venham visitá-lo pela primeira vez, sem qualquer conhecimento técnico, com o intuito de introduzir a Visualização de Informação ao público geral.

⁶ Disponível em: <http://manyeyes.alphaworks.ibm.com/manyeyes/>

2.2.2 PERCEPTUAL EDGE

Stephen Few⁷ é um dos investigadores mais activos da Visualização de Informação, Focalizando-se na área de Visualização na análise de dados e comunicação de informação empresarial.

A sua empresa, *Perceptual Edge*⁸, é uma empresa de consultadoria e formação de empresas sobre o *Design* de sistemas de análise de informação.

Escreve regularmente artigos, é uma presença assídua em conferências, e dá aulas no *MBA* da Universidade da Califórnia, em *Berkeley*. É uma das vozes mais objectivas e críticas na matéria, procurando distinguir boas e más práticas nos projectos de Visualização da actualidade, através de uma abordagem pragmática sobre se há ou não criação de conhecimento novo através destas ferramentas.

2.2.3 FLOWING DATA

Dirigindo-se ao público geral, este portal⁹ reúne artigos que apresentam dados estatísticos de interesse cultural sob a forma de Visualizações. Também descreve novas ferramentas de Visualização consoante a sua aplicação ou área temática. Tem como objectivo principal que toda a gente consiga compreender informação complexa relevante ao seu dia-a-dia através das visualizações.

O criador deste portal, *Nathan Yau*, é licenciado da Universidade da Califórnia, em Los Angeles (*UCLA*), e é um entusiasta da Visualização de dados social, ou seja, na promoção de grupos de indivíduos na análise e partilha de grandes conjuntos de dados; também se interessa na construção de Visualizações através de dados aparentemente inúteis e criados involuntariamente no dia-a-dia das pessoas.

2.2.4 VISUAL COMPLEXITY

Este portal¹⁰ aglomera diferentes métodos de visualização de Redes complexas, sobre várias áreas temáticas (Biologia, Redes Sociais, *WWW*). Os projectos apresentados no portal são seleccionados tendo em conta a originalidade do tema e a inovação das técnicas de Visualização empregues.

Lançado em Outubro de 2005, este site¹¹ surgiu para colmatar a falta de portais *on-line* focalizados na Visualização de Redes. Em complemento com a apresentação dos projectos, que podem ser comentados pelos utilizadores, o portal integra um *blog* que discute alguns tópicos sobre a Visualização de Informação e o Design de Interação nos sistemas actuais.

2.2.5 SPOTFIRE

Sendo um dos melhores exemplos da proliferação comercial dos sistemas de Visualização de Informação, o projecto *Spotfire*¹² foi inicialmente desenvolvido pela utilização de vários trabalhos sobre pesquisas dinâmicas e Campos de Estrelas para a gestão de Base de Dados, realizados principalmente por

⁷ É também autor de dois livros de referência: *Show me the Numbers: Designing Tables and Graphs to enlighten*, e *Information DashBoard Design the Effective Visual Communication of Data*.

⁸ Disponível em: <http://www.perceptualedge.com/>

⁹ Disponível em: <http://flowingdata.com/>

¹⁰ Disponível em: <http://www.visualcomplexity.com/vc/>

¹¹ O seu criador Manuel Lima, Designer de interação e Informação, trabalha actualmente na divisão dos Serviços e Design de Interação da *Nokia*. É também uma voz activa em muitas conferências sobre a Visualização de Informação, em especial sobre Redes complexas.

¹² Disponível em: <http://spotfire.tibco.com/>

*Ben Shneiderman*¹³ (Ahlberg, 2000). Foi das primeiras Visualizações de dados com a preocupação específica de atingir um público com pouca experiência na utilização destes sistemas, e rapidamente se transformou numa empresa com rápido crescimento.

Entre os seus clientes contam-se empresas de vários ramos, tais como a Química, Biotecnologia e Investigação. A CIA (Agência de Inteligência Central norte-americana) conta-se entre um dos seus investidores, considerando o potencial que pode existir ao comparar por exemplo milhares de casos de tribunal (Datta, 2007).

Um exemplo dos seus produtos, o *DecisionSite*TM permite aos utilizadores introduzirem e misturarem dados de vários tipos através da manipulação directa de imagens e ferramentas analíticas (Chen, 2006). Este produto tem também versões personalizáveis, consoante a empresa cliente.

2.2.6 TABLEAU

Análoga à empresa anterior, a *Tableau Software*¹⁴ é uma empresa que segue premissas semelhantes, mantendo o mesmo *slogan* de criar ferramentas de visualização para toda a gente, mas pretende atingir principalmente entidades empresariais como cliente. Começou como um projecto do Departamento de Defesa Norte-americano, liderado por *Pat Hanrahan*, professor da Universidade de *Stanford*, na Califórnia. A equipa desenvolveu a tecnologia *VizQL*TM, uma linguagem formal para descrição de objectos (mapas, tabelas e gráficos), que permite a rápida transição entre diferentes estilos de Visualizações (Hanrahan, 2006).

Ganhou o estatuto do Produto do Ano de 2005, atribuído pela revista *PC Magazine*¹⁵.

2.2.7 XPLANE

Esta empresa¹⁶, fundada em 1993, dedica-se à consultadoria de Design de Informação, procurando estratégias visuais de raciocínio e comunicação da informação, com o intuito de resolver vários problemas comunicacionais das empresas clientes; estes problemas dependem exclusivamente do processo de negociação; podem ter diferentes objectivos: o melhoramento das relações inter-divisões da empresa, através do mapeamento do sistema comunicativo da empresa, o auxílio do lançamento de um novo produto, através da visualização da sua proposta e promoção; a formação de investidores, através de suportes didácticos visuais.

Dave Gray lidera esta empresa, sendo uma voz activa sobre formas de pensar e trabalhar visualmente, estando presente em várias conferências e workshops na área da Visualização de Informação. É o fundador da comunidade global online *VizThink*¹⁷, que organiza debates, conferências e workshops pelo globo, sobre o pensamento visual. Esteve bastante empenhado na formação a educadores, corporações e ao público no geral, sobre a comunicação visual de informação.

¹³ Sendo originalmente fundada em 1996, a empresa composta de cerca de 200 profissionais, foi comprada em 2007 pela empresa TIBCO, quando afirmava obter um crescimento médio anual de 30% desde a sua fundação.

¹⁴ Disponível em: <http://www.tableausoftware.com/>

¹⁵ . Afirma que tem actualmente mais de 10.000 clientes pelo mundo inteiro (que ainda representa um volume de negócios mais reduzido que a Spotfire, que é uma empresa mais antiga e mais integrada no mercado).

¹⁶ Disponível em: <http://www.xplane.com/>

¹⁷ Disponível em: <http://www.vizthink.com/>

2.2.8 GAPMINDER

Merece também especial destaque esta organização¹⁸ com fins não lucrativos, que tem como objectivo o esclarecimento e divulgação de dados estatísticos sobre o desenvolvimento social, económico, e ambiental dos diferentes países, numa escala global.

Esta iniciativa surge para combater os preconceitos existentes ainda sobre diferentes países, tais como associações do tipo “Na África todos os países são pobres e com muitas doenças”, ou “No mundo temos os países ocidentais e os países do terceiro mundo”.

Através de aplicações de Visualização interactivas, a equipa do *Gapminder* propõe-se a tornar esta informação, que já existe disponível, fácil de perceber e divertida. Nas suas conferências, *Hans Rosling*, um dos fundadores, consegue tornar os gráficos de bolhas (*Bubblecharts*) muito elucidativos e divertidos, enquanto vão retratando o desenvolvimento dos países do globo ao longo dos anos até hoje.

Este projecto é um grande exemplo de como se consegue transmitir com sucesso ao público geral, uma maior compreensão do panorama socioeconómico global, através de visualizações relativamente simples.

2.3 PERCEPÇÃO E COGNIÇÃO

No que diz respeito aos mecanismos cognitivos que são desencadeados pela Visualização de Informação, é relevante considerar não só os mecanismos do corpo humano que reguem os estímulos visuais que recebemos do mundo exterior, mas também a maneira como o cérebro os interpreta (Ware, 2004).

Deste modo, é necessário perceber de que forma é que se podem criar estímulos visuais representativos de algo. Da mesma forma é também essencial compreender como se destaca um item, ou um grupo de itens específicos dentro de uma multiplicidade de sinais exteriores.

Colin Ware (2004) é um investigador que no seu estudo da Visualização de Informação se centrou principalmente no lado da percepção humana, e dos sistemas do cérebro que dela fazem parte. O seu trabalho continua a ser o mais extenso nesta área, portanto seguimos alguns pontos apresentados por ele nesta investigação.

Como ponto de partida, convém perceber que o que realmente estamos a ver numa visualização é um conjunto de símbolos.

Os símbolos são no fundo as unidades representativas do real, racionalizando informação abstracta em informação convencional, e podem ser divididas em duas categorias – Arbitrária e Sensorial (Ware, 2004).

Os símbolos que se designam de arbitrários precisam de uma aprendizagem prévia, pois a relação que mantêm com os objectos reais que representam é maioritariamente convencional, pré-estabelecido por algo. São uma construção social, cuja maior parte da sua influência deriva da cultura onde estão inseridos. São a base de qualquer raciocínio complexo, desempenhando o papel de unidades construtoras de qualquer linguagem formal, tal como a matemática ou a língua portuguesa. Como tal são difíceis de aprender, na medida em que precisam de uma aprendizagem rigorosa, tal como quando se aprende uma linguagem nova. Acresce ainda o facto de que estes símbolos podem ser facilmente esquecidos, se não forem utilizados regularmente (Ware, 2004).

Os símbolos Sensoriais são os símbolos que não precisam de um manual de instruções para se fazerem compreender. Dependem do processamento perceptual do cérebro que não precisa de aprendizagem prévia. A sua apreensão funciona de um modo inconsciente, porque o cérebro já está pré-programado para

¹⁸ Disponível em: <http://www.gapminder.org/>

os detectar. Isto faz com que não estejam presos à cultura do indivíduo, pois não são fruto de nenhum tipo de convenções conceptuais, o que faz com que sejam compreendidos em qualquer lado (Ware, 2004).

Na verdade, poucas vezes um símbolo é exclusivamente do tipo arbitrário ou do tipo sensorial. O que acontece normalmente é que reúne mais características de um deles, sendo a sua própria categorização um processo difícil e subjectivo. Tal facto não altera a importância que existe em perceber as diferenças entre os dois, pois estes têm aplicações e resultados diferentes na sua visualização (Ware, 2004).

Uma aplicação que fosse inconscientemente perceptível para toda a gente seria algo a aspirar pelos *designers* de comunicação: não impunha ao utilizador qualquer tipo de aprendizagem prévia para que este conseguisse utilizá-la correctamente.

Porém, na realidade os símbolos arbitrários têm de estar de alguma forma presentes em qualquer Visualização, sendo essenciais para a execução de análises mais complexas e detalhadas da informação (legendas, rótulos, dados estatísticos).

Esta investigação pretende no entanto focalizar-se mais nos símbolos predominantemente sensoriais; sendo universais, não estão dependentes de um contexto de uso nem do tipo de dados, podendo os seus conceitos serem usados em qualquer projecto de Visualização. É através da sua exploração que se devem basear as metáforas visuais, necessárias para uma Visualização intuitiva de dados abstractos.

2.3.1 PERCEPÇÃO PRÉ-CONSCIENTE

Pretendendo usufruir desta característica inconsciente dos símbolos sensoriais, existe uma série de medidas que conseguem destacar elementos dentro de uma visualização de uma forma imediata.

Certas diferenciações de formato podem fazer com que um objecto “salte à vista” do contexto onde está inserido. Tal é particularmente importante para as Visualizações de dados; muitas vezes o utilizador tem a necessidade de identificar instantaneamente uma marca particular dentro de um mapa ou de agrupar marcas semelhantes das outras.

A chamada Percepção Pré-consciente ocorre antes de qualquer tomada de atenção consciente, e determina quais os objectos merecedores da nossa imediata atenção.

Através de várias experiências realizadas, com o intuito de identificar quais os tipos de imagens que mais se destacavam antes da percepção consciente, conseguiu-se criar uma categorização para estes estímulos. Ware divide os elementos de destaque pré-consciente nas seguintes dimensões: Cor, Forma, Posição Espacial e Movimento (Ware, 2004).

Jacques Bertin, no seu conhecido trabalho *Semiologie Graphique*, descreve seis variáveis gráficas, de onde Ware se baseia e complementa, através destas dimensões que define.

Para além disso, Bertin estabelece uma relação entre o tipo de dados e as dimensões visuais utilizadas: de um modo geral, nomeia o tamanho como apropriado para representar dados numéricos, e a tonalidade da cor, a forma e a orientação como mais adequadas para dados nominais (Bertin, 1983).

a) COR

Em exemplos mais práticos, imaginemos que temos um emaranhado de dados onde alguns deles têm valores fora da norma, e que precisem de atenção imediata - um mapa do país, onde temos uma distribuição por distrito dos riscos de incêndio; neste caso é usual verificar-se a diferenciação por distrito através de uma escala de cores, referente a esse perigo – as zonas de cor mais acentuada serão as cores que normalmente merecem mais atenção.

Existem 3 categorias da cor que podem funcionar de forma autónoma como elemento diferenciador do meio: tonalidade, saturação, e luminosidade (Ware, 2004).

A tonalidade diz respeito à posição da cor no espectro da luz visível, e está dependente do seu comprimento de onda correspondente; (Ware, 2004) de certo modo é a característica que nos indica se um objecto é vermelho, azul, amarelo, por exemplo. Imaginemos uma bóia vermelha à deriva no mar: esta destacar-se-á mais facilmente devido ao contraste que existe entre a sua cor vermelha e o azul do oceano.

A saturação está ligada à intensidade da cor, se está muito *forte* ou *fraca*; por outras palavras indica a medida de pureza da cor, na medida em que quanto mais pura for a sua tonalidade, mais saturada estará a cor e vice-versa (Ware, 2004). Imaginemos bóias vermelhas que estão empilhadas umas em cima das outras, as bóias novas destacar-se-ão das bóias velhas, pois têm a cor menos gasta – têm uma maior saturação da mesma cor vermelha.

A luminosidade traduz o brilho que determinada cor aparenta ter, que está ligado à quantidade de cor branca que está “diluída” nessa cor: quanto maior for a luminosidade de um objecto, mais embranquecida vai-se tornando a sua cor, até ao ponto de se tornar imperceptível; o mesmo sucede no caso oposto: quando a luminosidade é muito reduzida, a cor desvanece-se. De noite, uma bóia está à deriva no oceano, banhada apenas pela luz da lua - provavelmente destacar-se-á melhor se tiver a cor amarela, pois esta apresenta uma maior luminosidade, do que se tivesse a cor vermelha.

Na Visualização de Informação, é comum utilizar-se a conjugação destes elementos para diferenciar os objectos pela cor. No entanto, pode dar-se o caso, por ex., em que temos grupos de dados agregados, cada um com a sua cor correspondente, e queremos dentro de cada grupo diferenciar elementos pontuais; de forma a não ficarmos com uma mescla de tonalidades na mesma visualização, podemos optar pela luminosidade ou pela saturação com diferenciadores, que possam destacar esses elementos de forma subtil mas perceptível.

b) FORMA

Noutro caso, imaginemos um cenário de uma rede social, onde se pretendem destacar os elementos com maior número de ligações bidireccionais (de entrada e de saída) com os outros elementos da rede; aqui podemos simplesmente aumentar o tamanho da representação desses nós, fazendo com que tenham maior relevância no panorama geral da visualização.

A dimensão da Forma engloba, não só o tamanho, mas também outros elementos tais como a orientação, a forma da figura, a focagem e a desfocagem, o agrupamento espacial e a curvatura (Ware, 2004).

Num contexto mais abrangente, a noção da Forma na Visualização pode implicar várias noções complexas de *Design*, sobre representação icónica de informação. No entanto dentro do contexto da percepção pré-consciente, não faz sentido aprofundar muito esta ideia, pois a percepção de um ícone representativo envolve processos de associação mental mais complexos, muitas vezes dependentes de um factor cultural, enquadrando-se já na categoria de símbolos arbitrários.

O que importa salientar aqui é que dentro de um conjunto homogéneo de formas, qualquer forma diferente do grupo é a mais saliente.

c) MOVIMENTO

Também se podem diferenciar elementos com base da sua informação cinética. Os elementos que se movimentam “contra a corrente” são os que ressaltam mais à vista, porque aparentam fugir à norma. Num cenário em que temos um fundo estático, é simples realçar determinados objectos através da sua vibração, ou brilho animado, por ex., porque estes são os únicos que se movimentam; quando existem vários elementos animados em cena, é preciso ter em conta outros factores, tais como a orientação do movimento (Ware, 2004).

Este tipo de percepção também é contextual, no sentido em que a movimentação dos elementos vai ser percebida dependendo do movimento do meio onde estão inseridos; como exemplo prático temos uma figura estática inserida numa moldura em movimento lateral, com um plano de fundo homogéneo; neste caso temos a percepção de que é a figura que se está a mexer, em vez da moldura (Ware, 2004).

Existe todo um conjunto de aplicações mais óbvias para uso do movimento nas Visualizações: a representação de canais de comunicação ou da dinâmica de partículas. Porém existem outras pequenas aplicações discretas que podem ter um impacto perceptual. Na verdade, paralelamente ao conjunto das dimensões estáticas, existe todo um vocabulário de movimentos expressivos capazes de comunicar significado; através do movimento é possível reconhecer directamente relações de causa e consequência, assim como as relações directas entre dados complexos; tal como referido anteriormente, as animações não necessitam de serem representações elaboradas; os detalhes complexos deixam de ser necessários podendo até ser factores de distração.

d) Posição

Recorrendo ao exemplo da visualização de uma rede social, poderíamos utilizar a posição espacial dos nós da rede para descrever o grau de proximidade destes - os nós espacialmente próximos denominariam grupos distintos de elementos.

A posição espacial numa representação virtual a 3 dimensões utiliza exactamente os mesmos propósitos, com a inovação de ser adicionada a dimensão da profundidade.

Quando não dispomos de uma visualização a 3 dimensões podem-se utilizar conceitos de perspectiva para dar a ilusão de existir um espaço tridimensional imaginário onde se distribuem os elementos da visualização; ferramentas que transmitam perspectiva podem ter uma implementação técnica bastante simples, permanecendo eficazes.

A sobreposição de elementos diferentes sugere automaticamente a existência de profundidade na visualização; se o objecto B é parcialmente ocultado pelo objecto A, temos a sensação que B está atrás de A (Ware, 2004).

A focagem de um objecto pode sugerir que está mais próximo do utilizador em relação ao que está desfocado, atribuindo-lhe uma aproximação em profundidade; aqui estamos a simular o nosso comportamento no mundo real: focamos em determinados objectos que nos captam interesse e tudo o resto fica desfocado em segundo plano (Kosara, Miksch & Hauser, 2002).

A projecção de sombras sobre um objecto também consegue atribuir-lhe o efeito de destaque, uma vez que esta projecção faz com que se perceba que o objecto está mais próximo do que o plano invisível onde esta é visualizada. Esta projecção de sombras é especialmente forte quando utilizada em objectos em movimento (Kersten, Mamassian & Knill, 1997).

Também se recorre muitas vezes a mecanismos contextuais que sirvam como referência para uma representação a 3 dimensões, podendo surgir sob a forma de um simples referencial de coordenadas, de um “nevoeiro” simulado que esbate os objectos mais afastados.

O recurso à posição espacial como elemento de destaque detém um estatuto privilegiado no universo da Visualização de dados; a maior parte dos utilizadores já aprenderam a metáfora da navegação espacial no uso de ferramentas, desde na análise de gráficos de funções matemáticas à utilização de sistemas operativos de computador; tal conhecimento prévio atenua significativamente a curva de aprendizagem na utilização de novos sistemas de informação que seguem estas metáforas (Ware, 2004).

É comum conjugarem-se várias dimensões para destacar o mesmo elemento do resto. Tal pode estar relacionado com questões de acessibilidade – não basta distinguir um elemento pela cor, quando existe o perigo do utilizador ser daltónico, por ex.

Um exemplo popular que usa um paradigma semelhante são as chamadas *tag clouds*, popularmente utilizadas nos *sítios web* (*blogs*, portais, comunidades) para mapear os seus tópicos mais procurados. Isto acontece pela visualização das *tags*, que são categorias definidas no portal, associadas a categorias de conteúdos. Este mapa permite discernir rapidamente quais são os tópicos que mais se destacam, com um maior número de leitores, mas pode também ser uma ferramenta de navegação, quando cada *tag* representada funciona como *hiperligação* para o conteúdo associado a ela.

Do ponto de vista da utilidade cognitiva, discute-se se as *tag clouds* acrescentam algo de novo como Visualizações, no sentido em que não criam novo conhecimento, sendo facilmente substituídas por tabelas de dupla entrada, que mostram os dados de modo mais claro e organizado. A distribuição das *tags* não segue de facto nenhum critério de proximidade semântica; não representam o grau de proximidade entre os seus elementos, um factor que poderia ser interessante para a análise das relações semânticas estabelecidas entre si. Por outro lado as *tags* que tenham uma contagem reduzida de ocorrências, que não deixam de ser indicadores importantes, têm tamanho excessivamente diminuído, sendo muitas vezes apenas “ruído de fundo” imperceptível (Hearst, 2008).

No que toca a dinâmica da procura de conteúdo das *tag clouds* – o algoritmo de contagem sucessiva dos cliques não é suficiente para representar os interesses dos utilizadores. Estes cliques podem significar muitas vezes “visitas relâmpago”, em que o utilizador entra na página mas sai logo a seguir, por não ter nenhum interesse no conteúdo que encontrou. Devia existir uma representação dessa contagem ao longo do tempo. Através de uma análise ao longo do tempo, podemos ter uma maior aproximação das tendências do utilizador.

No seu trabalho de investigação, Moritz Stefaner (Stefaner, 2007) explora novos modelos de visualização para tentar colmatar as limitações das *tag clouds* actuais²⁰.

Um dos modelos utiliza a elasticidade como factor de interacção (Figura 6). Neste caso, o principal objectivo é mapear a forma como as *tags* surgem em conjunto, ao revelar estruturas emergentes compostas de *tags* que estão semanticamente relacionadas. Cada *tag* é escalado de acordo com a sua frequência (como sucede na *tag cloud* típica) mas está também ligado semanticamente com outras *tags* presentes na amostra. A elasticidade traduz-se na capacidade de arrastar as *tags*, cujos movimentos também inferem nas restantes *tags* que lhe estão ligadas. A partir do clique e arraste, as *tags* seleccionadas pelo utilizador podem ficar na posição espacial que este entender, separadas da restante nuvem. Por outro lado, quando se voltam a clicar nestas *tags* destacadas, estas voltam para a sua disposição espacial dentro da nuvem geral. Este movimento é acompanhado pela alteração da “elasticidade” das suas relações com as outras *tags*.

O autor afirma que esta Visualização por si só não é suficiente: perde-se a dimensão abstracta dos conceitos ao representa-los espacialmente com um “corpo” concreto, confinado a um lugar “físico”. A metáfora visual enfatiza demasiado na distância entre os elementos, não estando esta verdadeiramente ligada a nenhum indicador, sendo muitas vezes atribuída aleatoriamente.

²⁰ Disponível em: http://well-formed-data.net/experiments/tag_maps_v5/

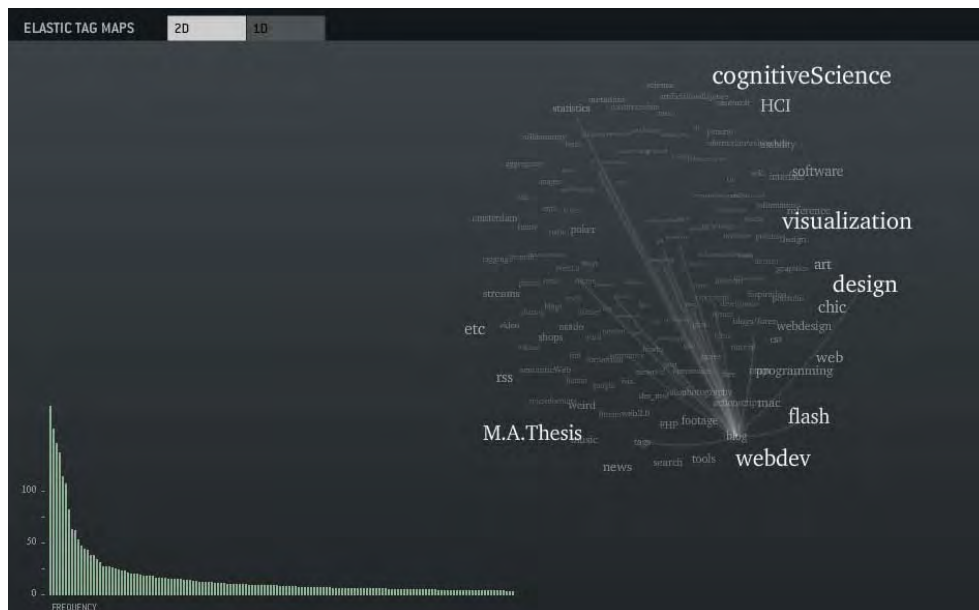


FIGURA 6 - ELASTIC TAG MAPS²¹, DE MORITZ STEFANER, ONDE DO LADO DIREITO VEMOS A TAG CLOUD ELÁSTICA, COM ALGUNS ELEMENTOS EM DESTAQUE, E DO LADO INFERIOR ESQUERDO, O HISTOGRAMA RELATIVO ÀS FREQUÊNCIAS DAS PALAVRAS.

Para contrapor estes resultados, a aplicação é complementada com o paradigma de lista, passando para uma representação de 1 dimensão (altura), mantendo alguns conceitos de interacção. Apesar de partilhar alguns problemas com a primeira experiência, esta abordagem permite na mesma ver elementos relacionados, de uma forma menos caótica. Esta modalidade representa um exemplo interessante da integração de modelos (de visualização e interacção) usados nas Visualizações actuais, em modelos clássicos pré-existentes, tais como tabelas e listas. Tal é particularmente interessante na medida em que se procura conjugar o carácter inovador e expressivo com a familiaridade e *know-how* que o público geral tem da análise de dados mais “tradicional”.

Outro modelo experimentado procura estudar a evolução da frequência das *tags* ao longo do tempo. É baseado no projecto *ThemeRiver™* (Havre, Hetzler, & Nowell, 1999). As *tags* são representadas por “rios”, dispostos ao longo do eixo horizontal. A grossura destes rios está directamente ligada à afluência de utilizadores a determinada *tag*. São acumulados no eixo vertical pela ordem da primeira ocorrência (quanto mais antigos, menor será a sua altura e quanto mais recentes, mais alto estarão), ou por frequência, conforme a escolha do utilizador.

O algoritmo utilizado encara as *tags* como algo que envelhece e é eventualmente esquecido ao longo do tempo, representando o seu ciclo de vida. É um protótipo particularmente útil se tivermos em conta a natureza dinâmica das tendências, em constante mudança – aqui é possível identificar o surgimento e desaparecimento das palavras-chave mais procuradas. É um exemplo que pode ser adaptado para a análise de outras tendências emergentes (tais como as palavras mais pesquisadas na *Web*).

Também é estudada uma versão animada da *tag cloud*, que representa a oscilação temporal das *tags*, através da variação do tamanho das palavras, e através do desaparecimento e aparecimento de novas *tags* na visualização. É uma modalidade que pode ser útil quando temos a necessidade de visualizar os dados em tempo-real e estudar tendências emergentes de uma forma mais imediata e transparente. Continua a ter problemas quando se pretende uma análise mais detalhada para a criação de conhecimento generalizante, na medida em que perdemos a noção da evolução, quando não temos pontos de referência com o passado (Stefaner, 2007).

²¹ Retirado de: http://well-formed-data.net/experiments/tag_maps_v5/

Existem também várias experiências que, tal como no projecto proposto para esta investigação, pretendem cruzar *tags* com referencias geográficas. Um exemplo desta junção é o projecto *Etsy Geolocator*²² (Weskamp, Etsy Geolocator, 2007) que através dum mapa a 3 dimensões do globo terrestre, permite ao utilizador localizar geograficamente locais de venda, *tags* de tipos de produtos, associados ao portal de venda e compra *online* de artigos artesanais.

Outras aplicações interessantes sobre a visualização textual funcionam como agregadores de *feeds*. Os agregadores de *feeds* recolhem automaticamente conteúdo novo publicado em páginas *Web* externas, permitindo ao utilizador ler este conteúdo directamente de uma única aplicação.



FIGURA 7 - DE UM AGREGADOR DE NOTÍCIAS EM FORMATO DE MAPA DE ÁRVORE: NEWSMAP²³

A aplicação *Newsmap*²⁴ (Weskamp, Newsmap, 2004) reúne as notícias do agregador *Google News*, distribuindo-as por temas, tais como Internacional, Negócios e Tecnologia. Para tal utiliza um mapa conceptual (Figura 7) que agrupa as notícias com temas semelhantes, utilizando uma tonalidade semelhante de cor. Também são destacados os tópicos que tenham aparecido mais vezes nas notícias, dispondo de títulos de maior dimensão e luminosidade cromática.

A metáfora visual que utiliza consiste no reaproveitamento do método de visualização conhecido como *Mapa de Árvore* para o paradigma textual, onde os cabeçalhos são as representações das notícias. É um trabalho que permite transmitir rapidamente uma ideia geral sobre quais os temas com maior movimentação e qual a variedade de tópicos existente.

Com objectivos semelhantes, o Projecto *Voyager*²⁵ (Biggs, 2008) adopta uma metáfora 3D para a ferramenta de agrupamento das notícias. Através de mecanismos de translação, de aproximação e de afastamento, é possível navegar por um espaço tridimensional, onde flutuam os cabeçalhos das notícias. Aqui o que se ganha em navegação perde-se em usabilidade. É possível que se torne mais difícil encontrar uma notícia específica, quando temos uma grande quantidade de *feeds* reunidos na aplicação.

Outro leitor de *Feeds* que se popularizou rapidamente devido à sua originalidade conceptual é o *We Feel Fine*²⁶ (Harris & Kamwar, 2006). Este projecto detecta numa janela temporal curta, textos escritos em vários

²² Disponível em: <http://www.etsv.com/geolocator.php>

²³ Retirado de: <http://www.cs.umd.edu/class/spring2005/cmssc838s/viz4all/ss/newsmap.png>

²⁴ Disponível em: <http://newsmap.jp/>

²⁵ Disponível em: <http://www.rssvovage.com/>

²⁶ Disponível em: <http://www.wefeelfine.org/>

blogs que tenham a conjugação de palavras “*I feel*” e “*I’m feeling*”. Após esta detecção, o sistema recolhe a frase completa onde estão inseridas estas palavras, procurando se nesta existe um dos 5000 sentimentos, enumerados e categorizados na base de dados do sistema (Figura 8). Os diferentes sentimentos são então representados visualmente pela metáfora de partículas, mas para outros tipos de Visualizações diferentes, que permitem recolher vários tipos de indicadores.



FIGURA 8 - O EXEMPLO DE UM DOS ECRÃS DA APLICAÇÃO *WE FEEL FINE*, ONDE NOS É APRESENTADA A LISTAGEM ORDENADA DOS SENTIMENTOS MAIS UTILIZADOS, DENTRO DOS PARÂMETROS ESTABELECIDOS²⁷.

2.5 PADRÕES DE DESIGN

Desde que a Visualização de Informação começou a emergir como área de investigação foram-se formulando novas propostas taxonómicas capazes de categorizar os seus mecanismos.

Shneiderman define que a exploração dos dados tende a seguir o seguinte caminho: Vista geral, focagem e filtragem, e detalhes à medida (*Overview first, zoom and filter, details-on-demand*) – *Mantra da Procura de Informação*.

Começarmos com uma visão geral de todo o conjunto de dados é muito importante, pois: *reduz a pesquisa a efectuar, permite a detecção de problemas gerais, e ajuda o utilizador a escolher o próximo passo* (Card, Mackinlay & Shneiderman, 1999).

Quando se identificam marcos visuais mais relevantes da visualização, é relevante focar um destes em particular, examinando-o mais de perto, o que implica a focagem em pontos específicos. A filtragem dos dados também é importante, no sentido em que se reduz o conjunto de dados para um tamanho mais manipulável, retirando o que não interessa. Os detalhes à medida (*details-on-demand*) consistem na examinação meticulosa dos dados, em que precisamos de saber informações adicionais.

A escolha destes detalhes deve ser simples e directa, disponível apenas para quando precisarmos delas. Desta forma elas não estão sempre no caminho, interferindo com a visualização como ruído visual (Few, 2006).

Christian Behrens propõe uma taxonomia estrutural que categoriza o design de informação aplicado à área da Visualização de Informação. Para tal define um conjunto de padrões de Design, estipulados através

²⁷ Retirada de: <http://aesthetechtonik.files.wordpress.com/2007/06/we-feel-fine-01.jpg>

da análise de diferentes sistemas de visualização da actualidade. O trabalho procura criar uma base estrutural, que esteja em constante actualização, procurando acompanhar o surgimento de novas ferramentas (Behrens, 2008).

Esta taxonomia é particularmente relevante no âmbito desta investigação, porque é um dos mais recentes estudos efectuados, desenvolvido com base nas tipologias populares de *Shneiderman* (Card, Mackinlay & Shneiderman, 1999). Procura exemplificar de forma objectiva e estruturada várias soluções de Visualização actuais, algo que ainda não foi muito bem aprofundado. Por outro lado, visa atingir especificamente os designers gráficos que estão a trabalhar nesta área que tenham pouca (ou nenhuma) experiência estatística e científica.

Não ambiciona servir de manual de instruções, que prescreva rigorosamente quais os estilos a adoptar para cada problema; em vez disso, procurar apoiar as escolhas do designer, fornecendo-lhe pistas sobre as representações visuais que funcionam melhor consoante o tipo dos dados em questão.

Os padrões definidos diferenciam-se em 3 categorias principais: Padrões de Visualização, Padrões de Comportamento e Padrões de Interação (Behrens, 2008).

Os padrões de Visualização (*Display Patterns*) centram-se em questões de como traduzir visualmente a informação, tendo em conta o tipo de dados e a sua aplicação pretendida.

Por sua vez, os padrões de Comportamento (*Behavior Patterns*) lidam com as medidas mais usadas pelo utilizador ao efectuar a análise de dados nas Visualizações.

Finalmente, os padrões de Interação (*Interaction Patterns*) também estão ligados à interactividade da visualização, centrando-se mais nos próprios elementos de interface que permitem a manipulação directa dos dados.

2.5.1 PADRÕES DE VISUALIZAÇÃO

Estes padrões estão relacionados à forma como nos são apresentados e organizados os dados, com o intuito de criar novo conhecimento. Estão intimamente ligados aos tipos de dados que vão ser consumidos, mas também aos objectivos de análise para estes dados e ao tipo de perguntas que pretendem responder. Com base nestas características, estes padrões dividem-se em: Correlações, Quantidades contínuas, Quantidades discretas, Proporções, Fluxos, Hierarquias, Redes e Configurações espaciais (Behrens, 2008). Por cada uma destas subdivisões enumeram-se alguns exemplos de ferramentas mais usadas.

a) CORRELAÇÕES

As correlações dizem respeito à relação directa entre duas ou mais séries de dados quantitativos, através da correspondência existente em cada par de variáveis (Behrens, 2008). Esta distribuição é dada normalmente dentro de um eixo bidimensional, cujas dimensões cartesianas correspondem a cada série de dados.

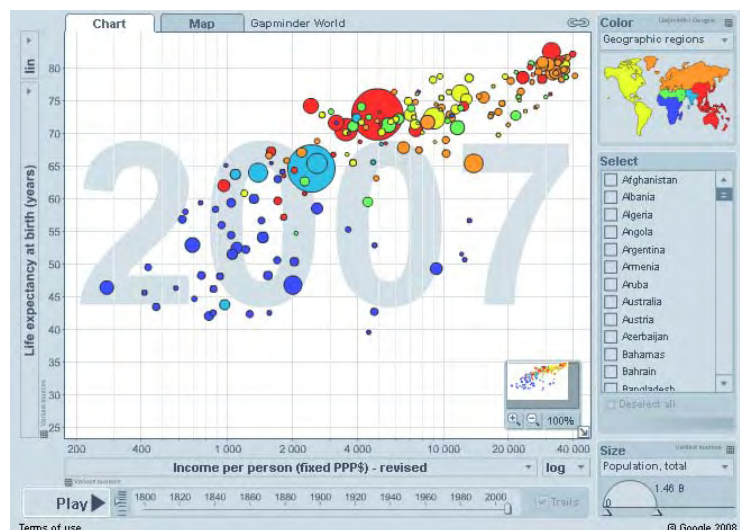


FIGURA 9 - EXEMPLO DE UM BUBBLECHART, UMA DAS FERRAMENTAS DE VISUALIZAÇÃO DO GAPMINDER²⁸.

Quando temos dois conjuntos de dados a correlacionar, é frequente o uso das matrizes de dispersão (*Scatterplot*). Consistem num gráfico bidimensional que apresenta através de pontos (ou outro tipo de símbolos), as relações existentes entre os dados de duas séries, que seriam imperceptíveis numa tabela de dupla entrada (Behrens, 2008). É possível representar a relação simultânea entre pares de grupos de dados, através de diferentes escalas de cor, por ex.

Por outro lado, podem-se representar mais do que duas dimensões de dados, dentro do mesmo sistema bidimensional de coordenadas; atribui-se o valor da terceira dimensão de dados à representação visual do próprio objecto - que deixa de ser um simples ponto; os gráficos de bolhas (*Bubblechart*), usam o raio circular das círculos como indicador desse terceiro valor - quanto maior o valor, maior será o tamanho da bolha (Figura 9). É possível representar mais do que 3 dimensões, se acrescentarmos, por ex., a tonalidade da cor como factor de diferenciação extra, entre outros (Behrens, 2008).

b) QUANTIDADES CONTÍNUAS

Muitas vezes, é importante estudar a evolução temporal de determinados grupos de dados, cujos valores variam de forma contínua. Apesar destes dados serem usualmente resultado de uma amostragem discreta, podem ser tratados na mesma como uma linha contínua, porque têm uma frequência de amostragem elevada (Behrens, 2008).

As suas ferramentas mais comuns consistem em gráficos de duas dimensões onde um dos seus eixos de coordenadas (vertical ou horizontal) serve de eixo temporal, constituído por intervalos (de tempo) regulares. Os gráficos de linhas (*Simple Line Chart*) são representados através de uma linha contínua que se altera ao longo do tempo. Estes gráficos diferenciam-se assim das correlações, cujos valores representados são pontuais e discretos (Behrens, 2008).

Complementarmente, os gráficos multi-linhas (*Multiset Line Charts*) agregam vários conjuntos de dados no mesmo gráfico, através de várias linhas, correspondentes a um conjunto distinto de dados em função da mesma escala temporal. Normalmente utilizam-se diferentes tonalidades de cor se para diferenciarem as diferentes linhas (Behrens, 2008).

Estas ferramentas são úteis para comparações de diferentes evoluções temporais de forma separada; quando se pretendem comparar séries de dados dentro de um determinado contexto, o Gráfico de Áreas Empilhadas (*Stacked Area Chart*) é mais adequado. Aqui o que se pretende é perceber qual o nível de

²⁸ Retirado de: <http://graphs.gapminder.org/world/>

importância das diferentes variáveis, tendo em conta a totalidade dos dados. A construção deste gráfico consiste numa série de procedimentos: desenha-se a primeira série numa área de determinada cor; a construção da próxima área soma os valores da série anterior, de forma a ficar sempre superior a esta, e assim sucessivamente. Assim, vão-se empilhando as áreas umas sobre as outras; a altura de cada série dentro de um intervalo de tempo representará o peso desta em relação às outras nesse intervalo de tempo (Behrens, 2008).

A aplicação *ThemeRiver™* (Havre, Hetzler & Nowell, 1999), referida anteriormente, vem elaborar mais esta tipologia. Ao acrescentar um factor de distorção no gráfico, acrescenta mais fluidez e beleza estética à visualização. Esta tipologia, agora denominada *Stream Graph*, parece ser cada vez mais utilizada para representar alterações ao longo do tempo (Figura 10). Propõe-se sobretudo para contrapor a ilegibilidade dos Gráficos de Áreas Empilhadas que se verifica quando se lida com um grande número de séries de dados, onde as séries com menor dimensão tornam-se difíceis de detectar (Byron & Wattenberg, 2008).

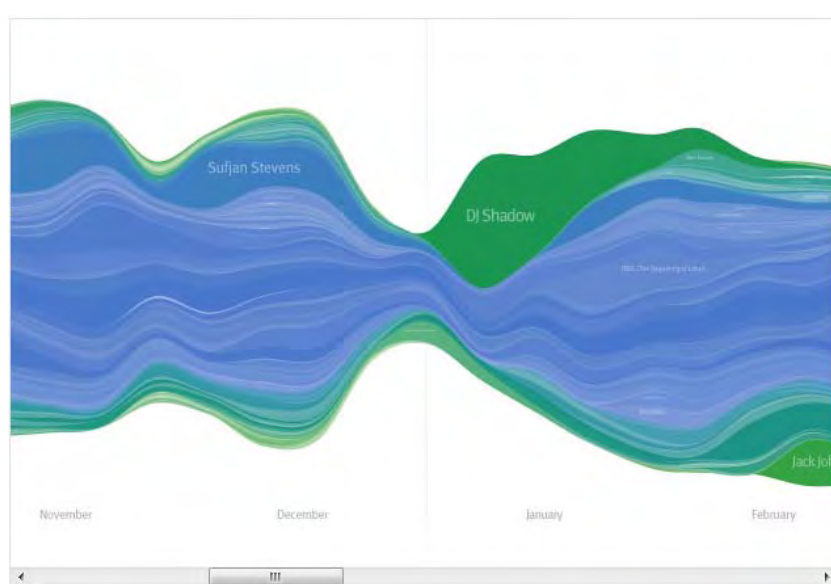


FIGURA 10 - LISTENING HISTORY – LAST.FM²⁹, UM EXEMPLO DE UMA APLICAÇÃO QUE USA A TIPOLOGIA DO STREAM GRAPH.

c) QUANTIDADES DISCRETAS

Em contraste com a secção anterior, aqui temos conjuntos descontínuos de dados, estruturados através de itens discretos de valores quantitativos. Aqui interessa saber concretamente as medidas de diferença existentes entre os valores, em detrimento de uma representação mais fluida e contínua (Behrens, 2008).

As formas mais comuns de representação são os gráficos simples de barras (*Simple Bar Chart*). São usados para comparar diferentes valores dentro do mesmo grupo de dados, mas existem variações que permitem a utilização de mais do que um grupo de dados. Os gráficos de barras multi-dados (*Multiset Bar Chart*) permitem a sobreposição de várias barras para cada variável da tabela de dados. Mais uma vez, é frequente o recurso à tonalidade, ou luminosidade da cor, para diferenciar os diferentes grupos de dados. De forma semelhante aos Gráfico de Áreas Empilhadas, o gráfico de barras empilhadas (*Stacked Bar Chart*) permite discernir o valor das diferentes componentes dentro das séries de dados em relação à sua totalidade (Behrens, 2008).

²⁹ Retirada de: <http://www.leebyron.com/what/lastfm/>

É possível criar gráficos que medem três dimensões de dados, através de um gráfico de barras isométrico (*Isometric Bar Chart*). Estes usam a projecção em perspectiva, criando um gráfico tridimensional, assente sobre um referencial bidimensional, onde a terceira dimensão é representada pela altura (Behrens, 2008).

As quantidades discretas também podem ser representadas de forma unidimensional. A matriz de pontos (*Dot Matrix*) retém a pontualidade dos valores, valorizando mais a contagem elemento a elemento; consiste numa matriz que lista visualmente elementos escalados representativos de um mesmo conjunto de dados, e diferenciados através da cor (Behrens, 2008).

Quando se analisam apenas os valores mínimos e máximos em séries de dados, o gráfico de extremos (*Span Chart*) é uma ferramenta simples e útil, composta de caixas cuja altura é definida por estes valores dentro de um sistema de coordenadas (Behrens, 2008).

O *Starplot* (Figura 11) por sua vez adequa-se a comparações de um número mais reduzido de variáveis de um mesmo elemento, sendo estas variáveis medidas através de um ponto comum (Spence, 2007).



FIGURA 11 - UM EXEMPLO DE UM *STARPLOT*³⁰, ONDE É UTILIZADA A COMPARAÇÃO DIRECTA ENTRE OS ATRIBUTOS DE DOIS ELEMENTOS (NESTE CASO, JOGADORES).

d) PROPORÇÕES

Este tipo de dados incide na comparação numérica das diferentes características, tendo em conta a totalidade do conjunto dos dados – normalmente sobre a forma de percentagem (Behrens, 2008).

O Diagrama de Área Polar de *Nightingale*, referido anteriormente, é o exemplo pioneiro da visualização deste tipo de relações. Hoje em dia os tipos de gráficos mais específicos e utilizados para representar proporções são os gráficos circulares (*Pie Charts*) e, derivados destes, os gráficos em anel (*Ring Charts*) (Behrens, 2008).

Os primeiros utilizam a divisão de um círculo segmentado, cujo tamanho dos segmentos (dado através da abertura do ângulo interno) é uma função directa da percentagem dos dados que representam, em relação à totalidade dos dados, representada pela totalidade do círculo, e são distinguidos normalmente por cores diferentes (Behrens, 2008).

Os gráficos em anel seguem a mesma metodologia, mas aplicada à representação de diferentes grupos de dados, e à comparação entre estes; constroem-se diferentes círculos concêntricos, relativos a cada série de dados, cuja posição dos segmentos pode estar alinhada propositadamente com outros segmentos de anéis diferentes, consoante a relação que tenham (Behrens, 2008).

³⁰ Retirado do jogo: Pro Evolution Soccer 2009

Existem várias outras ferramentas semelhantes aos dois exemplos apresentados, tais como o *SunBurst* (Stasko, Catrambone, Guzdial, & McDonald, 2000) ou a aplicação *DiskView*³¹, que são uma variação do mesmo, com inovações ao nível estético ou aplicacional, mas que mantém o mesmo paradigma estrutural.

e) FLUXOS

As representações que procuram descrever os fluxos têm como objectivo a compreensão das alterações dinâmicas que ocorrem sequencialmente num dado sistema (Behrens, 2008). Aqui interessa mais a evolução dos elementos, e de que forma estes se enquadram no sistema como um todo, usualmente suportado pelos seus valores percentuais.

Os Diagramas de *Sankey* (*Sankey Diagrams*) encaram a informação como um sistema complexo, de entradas e saídas, normalmente representado como um fluxo de energia. Este fluxo é dado muitas vezes sob a metáfora de setas, normalmente orientadas da esquerda para a direita, referentes aos dados de entrada e de saída, respectivamente, ou pela convergência e divergência de fluxos de informação (Behrens, 2008).

Por outro lado, o Diagrama de Arcos (*Thread Arcs*) é mais usado para representar a relação sequencial dos elementos constituintes do sistema (Figura 12). Estes elementos são normalmente distribuídos ao longo de uma linha, e é através de arcos que se representam as ligações entre pares de elementos (Behrens, 2008).

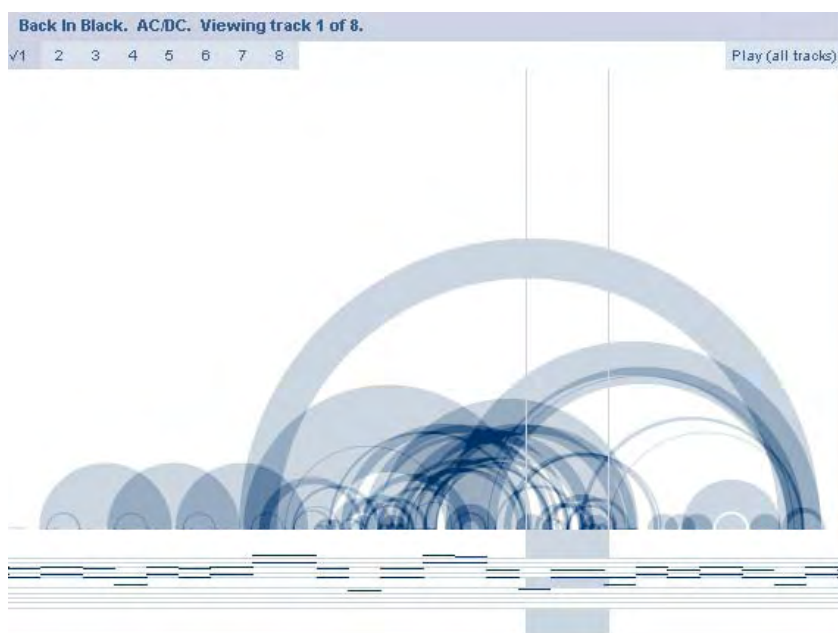


FIGURA 12 - A APLICAÇÃO *SHAPE OF A SONG*³², DE MARTIN WATTENBERG, PROCURA REPRESENTAR A FORMA DE CADA MÚSICA, MEDIDA ATRAVÉS DOS CICLOS MUSICAIS QUE SE REPETEM NA CANÇÃO, DESDE NOTAS A SEQUÊNCIAS MUSICAIS, REPRESENTADOS PELOS ARCOS.

f) HIERARQUIAS

Pode dar-se o caso de queremos estudar estruturas relacionais mais complexas, com um número considerável de componentes a interligarem-se heterogeneamente; estas ferramentas devem conseguir

³¹ Disponível em: <http://www.diskview.com/>

³² Disponível em: <http://www.turbulence.org/Works/song/gallery/gallery.html>

criar uma visão global do sistema, do qual transpareçam os diferentes níveis hierárquicos, mas também reflectindo características individuais dos elementos (Behrens, 2008).

Os Diagramas em Árvore (*Tree Diagrams*) distribuem espacialmente os elementos de um sistema, consoante a sua importância, e representam as ligações estabelecidas entre si (através de linhas); os elementos de maior destaque situam-se numa posição mais elevada (ou central) na estrutura e os elementos menos influentes situam-se numa posição menos elevada (ou mais periférica) da estrutura (Figura 13) (Behrens, 2008).

Através da metáfora de uma árvore, podemos identificar à partida quais os elementos com maior influência comunicativa; consegue-se discernir de que forma é que a raiz (elemento chave) chega aos ramos finais da hierarquia. É o caso da aplicação *Botanical Viewer*, que trás a metáfora de árvore para uma representação 3D de um modo mais literal (Kleiberg, Wetering & Wijk, 2001).

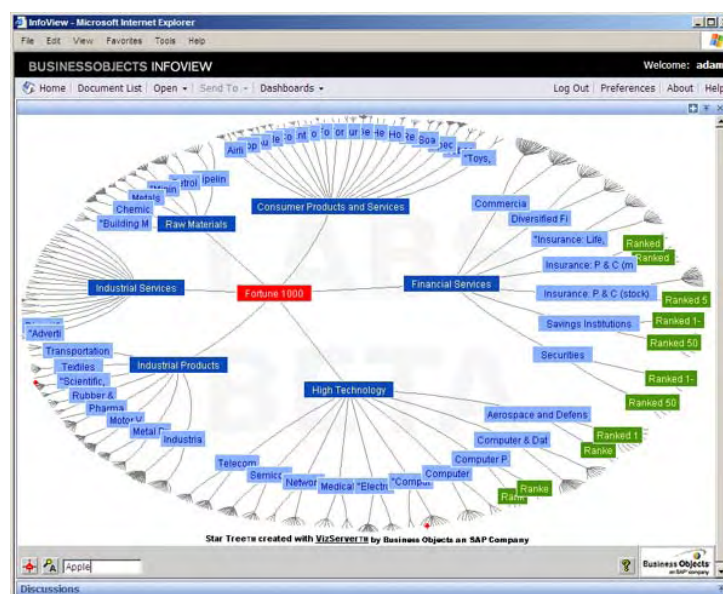


FIGURA 13 - EXEMPLO DE UM DIAGRAMA EM ESTRELA (*STARTREE*)³³, VARIANTE DO DIAGRAMA DE ÁRVORE, QUE ADOPTA UMA DISPOSIÇÃO CIRCULAR DOS ELEMENTOS, SENDO O ELEMENTO CENTRAL, O QUE OCUPA UMA POSIÇÃO DE TOPO NA HIERARQUIA.

O Mapa de Árvore (*Treemap*) é mais um modelo com a mesma tipologia, mas procura ter também em conta as características quantitativas dos elementos, ao mesmo tempo que mantém a relação hierárquica. Aqui os elementos são distribuídos na mesma de acordo com a sua posição dentro da hierarquia, mas o seu grau de importância é representado através de atributos tais como tamanho e cor. A posição espacial dos elementos passa a ser usada para agrupar elementos com categorias semelhantes (Behrens, 2008). Em norma, este método é escolhido quando se pretende uma rápida detecção de valores extremos na hierarquia. Um bom exemplo deste sistema é a aplicação *Map of the Market*³⁴ (referido mais à frente, na secção *Usabilidade*).

g) REDES

Quando não existe à primeira vista uma hierarquia identificável nas relações entre os elementos, interessa mais inferir nas suas ligações, que podem ter um número considerável (Behrens, 2008). Estas ligações vão criando percursos comunicativos, importantes para as representações de redes de transportes, assim como podem descrever o grau de conectividade de uma rede social *Web*.

³³ Disponível em: <http://labs.businessobjects.com/vizserver/default.asp>

³⁴ Disponível em: <http://www.smartmoney.com/map-of-the-market/>

O Diagrama de Mapa (*Diagram Map*), introduzido por *Harry Beck* em 1930, continua bastante popular nas redes de transportes; consistem na representação simplificada de uma rede existente de vários nós, interligados entre si. As posições dos nós estão alinhadas segundo uma grelha estrutural, tentam permanecer próximos da sua posição geográfica actual. As relações são representadas por linhas regulares, tomando um número restrito de direcções (horizontais, verticais, e diagonais); o que interessa realmente é o ponto de partida e de chegada, quais as paragens que se devem ter em conta, em detrimento do percurso geográfico exacto que se vai efectuar (Behrens, 2008).

Outras aplicações enfatizam mais na conectividade entre os elementos, como é o caso do Circulo Relacional (*Relation Circle*), mais utilizados para descrever redes sociais, podendo também serem usados utilizando factores geográficos. Consiste na disposição circular dos diferentes itens da rede, onde por cada relação existente entre um par de itens desenha-se uma linha curvilínea. A verdadeira visualização vai-se surgir no interior do círculo: um complexo de linhas ligadas à sua periferia (Figura 14).

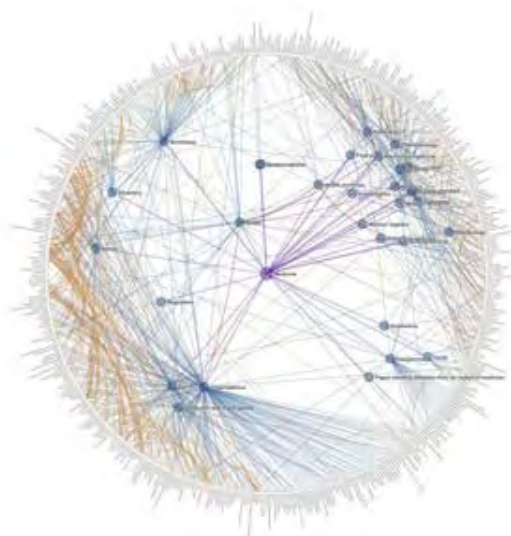


FIGURA 14 - EXEMPLO DE UM CÍRCULO DE RELAÇÕES – PROJECTO CLUSTERBALL: VISUALIZING WIKIPEDIA³⁵.

h) CONFIGURAÇÕES ESPACIAIS

Em complemento com o que foi discutido sobre a Geovisualização, encontramos aqui a descrição de alguns padrões comuns do *design* de informação que utilizam de algum modo dados georreferenciados. O que o autor designa de Mapas, como ferramentas de representar configurações espaciais são, na verdade, algumas das ferramentas utilizadas na Geovisualização (Behrens, 2008).

Dos mapas mais conhecidos, temos os Mapas Topográficos (*Topographic Maps*), que podem ter muitos formatos. Procuram agregar na mesma visualização tanto as características geográficas de uma localização real como as suas características culturais, tais como os nomes de ruas e restaurantes (Behrens, 2008). Os mapas variam muito nas suas características de Design, porque as suas especificações (nível do detalhe da área, tipos de elementos incluídos) variam com os objectivos concretos de cada projecto.

Por outro lado os Mapas Temáticos (*Thematic Maps*) concentram-se mais em valores quantitativos de séries de dados, e na sua relação geográfica. O mapa tem um carácter mais contextual e menos detalhado geograficamente, tendo o papel de englobar os conjuntos de dados sob uma entidade geocultural comum, tal como um país dividido por distritos.

³⁵ Disponível em: <http://www.chrisharrison.net/projects/clusterball/>

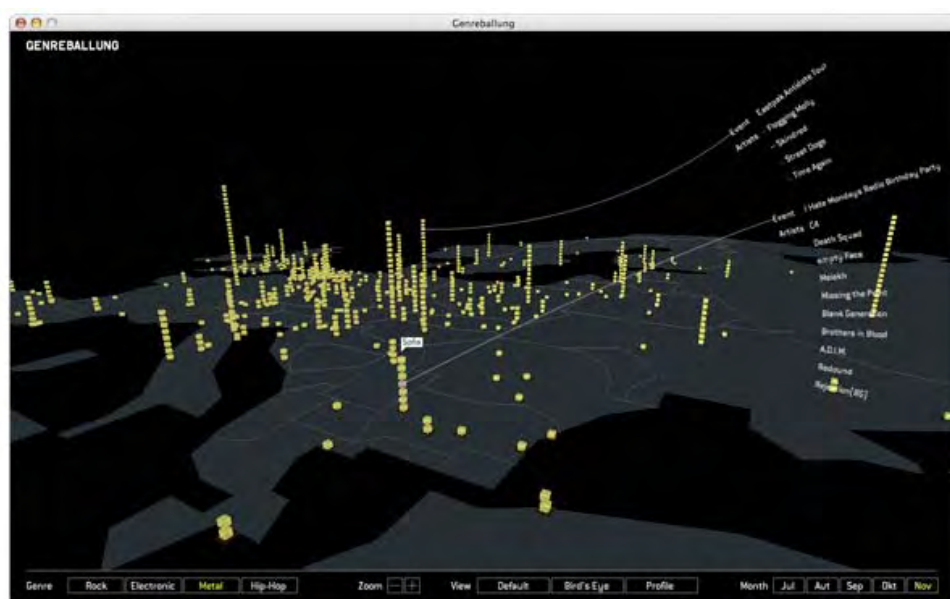


FIGURA 16 - APLICAÇÃO GENREBALLUNG, DE CHRISTOPHER ADJEI E NILS HOLLAND-CUNZ³⁹.

Tais exemplos são apenas uma amostra de todo um historial de aplicações que “fogem” do molde estrutural definido por *Behrens*; tal não é um facto particularmente antagónico ao seu trabalho, pois o principal objectivo desta taxonomia é distinguir os tipos de dados possíveis de serem representados e quais os tipos de visualizações mais específicos desses tipos de dados até à data.

Na verdade, é bastante difícil enquadrar todos os projectos de Visualização em algum tipo de categoria: como ainda estamos numa era de experimentação, novas metodologias vão surgindo. Interessa mais apreender a ideia geral sobre o que é preciso existir numa visualização esta se adequar ao tipo de dados que pretende analisar.

2.5.2 PADRÕES DE COMPORTAMENTO

Estes padrões dizem respeito às diferentes acções que são mais utilizadas para interagir com as Visualizações. Representam o carácter interactivo das interacções, que também estará ligado ao tipo de representação visual, mas sobretudo aos objectivos do utilizador para manipular os dados.

Estes padrões de comportamento estão agrupados por tipos distintos: Navegação, Filtragem, Arranjo, Exploração e Transição. Todos eles traduzem diferentes objectivos de interacção, produzindo resultados diferentes na Visualização (*Behrens*, 2008).

j) NAVEGAÇÃO

É frequentemente necessário ao utilizador ter um sentido de movimentação e de orientação dentro de uma Visualização. Este deve ter um sentido de direcção, saber para onde vai, o que procurar e deve conseguir fazê-lo sem esforço. Este sentido de percurso a caminhar começou com a metáfora dos mapas, mas rapidamente se estendeu para o resto das Visualizações (*Behrens*, 2008).

A função de *Zoom Simple* (*Simple Zoom*) possibilitam a rápida transição entre uma vista geral para uma vista particular localizada num excerto da informação e permitem ao utilizador definir o seu grau de

³⁹ Disponível em: <http://visualizinglastfm.de/Inhalt04/Metal1.gif>

proximidade com os dados. Complementarmente, o *Zoom Localizado (Local Zoom)* permite focar determinada área sem se perder o contexto geral – sendo a tradução literal de uma lupa (Behrens, 2008).

Quando temos uma Visualização de dados demasiado grande para se poder visualizar no ecrã de computador, utiliza-se frequentemente a opção de Translação (*Panning*); através das setas direccionais do teclado, ou arrastando o rato, para movimentar o mapa visual, de forma a posicionar a janela da Visualização sobre áreas que interessem.

Quando temos dados dispostos seguindo uma ordem cronológica, uma ferramenta bastante usual é a Linha temporal (*Timeline*). Permite navegar pelos dados consoante a janela temporal pretendida - logo estes dados devem ter uma dimensão temporal, que vai ser o objecto de interacção do utilizador (Behrens, 2008).

Quanto pretendemos a leitura de muitas variáveis que estão inter-relacionadas, torna-se contra-productivo não ser possível visualizar estas variáveis simultaneamente. *Linked multiples* refere a utilização de vistas Múltiplas, que implicam não só a visualização de várias janelas de dados, mas também mecanismos de interacção entre estas simultaneamente; estas janelas representam diferentes pontos de vista sobre a mesma informação, que pode levar a conclusões diferentes (Behrens, 2008).

A Vista Geral mais Detalhe (*Overview plus detail*), permite a selecção de uma secção do mapa geral, para visualizar dados detalhados dessa secção (Card, Mackinlay & Shneiderman, 1999). Através de uma miniatura do mapa geral, o utilizador selecciona determinada área, cujos pormenores transparecem no resto da visualização. Estando sempre presente esta miniatura permitirá ao utilizador saber onde está focalizado e saltar para outras partes.

k) FILTRAGEM

Esta secção agrupa métodos que incidem na selecção de parâmetros de entrada consoante a selecção do utilizador. Quando dispomos de dados complexos que albergam um grande conjunto de variáveis, a Visualização simultânea de todas essas variáveis torna-se ilegível. É então necessário filtrar o conteúdo, de forma a organizar a análise de dados em etapas, permitindo ao utilizador concentrar-se em dimensões compreensíveis de dados (Behrens, 2008).

Um mecanismo de filtragem bastante popular é a metáfora das Camadas (*Layering*) de informação, estando presente em todo o tipo de aplicações actuais. As camadas representam conceptualmente o amontoamento de informação sobre o mesmo objecto – tal como a sobreposição de películas transparentes sobre uma folha de papel, cada uma contendo informação distinta (Figura 17). Num ambiente virtual, estas camadas são tornadas visíveis pelo utilizador, através da sua activação ou desactivação (Behrens, 2008).

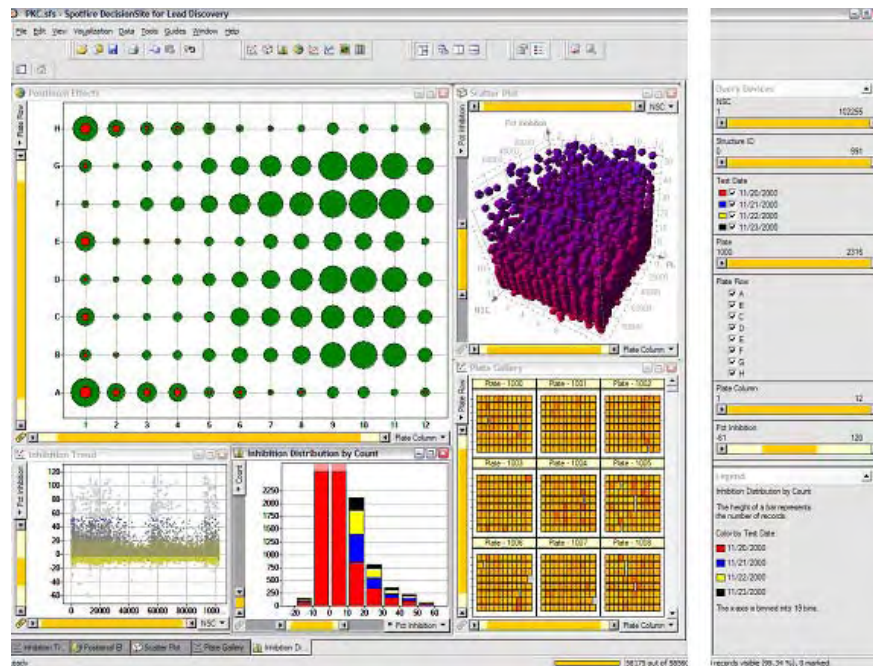


FIGURA 17 - UMA APLICAÇÃO DA EMPRESA SPOTFIRE⁴⁰, QUE CONJUGA VISTAS MÚLTIPAS (À ESQUERDA), MAS TAMBÉM CAMADAS DE INFORMAÇÃO, SELECIONADAS PELOS BOTÕES DA DIREITA.

É também possível utilizar o próprio carácter interactivo dos elementos para diferencia-los do resto. É este o propósito dos *Objectos Activos (Active Objects)* – tornar os objectos de destaque activos interactivamente, em contraste com o resto dos dados - que estão activos visualmente mas não interactivamente. Esta interactividade selectiva é normalmente visualizada através de factores cognitivos comuns, como cor, focagem e desfocagem (Behrens, 2008).

Também é possível recorrer a ferramentas de filtragem em série temporais, através da selecção do intervalo de tempo onde se pretende incidir o estudo - Filtros de Intervalos (*Boundary Filters*). Servem bem quando temos dados que possam ser limitados dentro de um intervalo, definido através de um valor temporal mínimo e máximo (Behrens, 2008).

As ferramentas de pesquisa continuam a ser uma forma válida de filtragem de informação. Porém, quando esta informação implica várias variáveis, uma pesquisa simples pode ser ineficaz. Uma Pesquisa Multi-facetada (*Facet Browsing*) permite personalizar o grau de especificidade da pesquisa, indicando quais as suas variáveis de entrada, possibilitando assim o cruzamento de categorias (Behrens, 2008).

Por outro lado a Pesquisa dinâmica (*Dynamic Query*) vai disponibilizando automaticamente sugestões de pesquisa e resultados preliminares enquanto o utilizador está ainda a indicar os parâmetros de pesquisa. Aqui procura-se auxiliar o utilizador, especialmente quando este tem dúvidas sobre o que está à procura, podendo dar pistas sobre o que procurar, para além de poupar tempo precioso de pesquisa, ao mostrar à partida se os resultados são os pretendidos ou não (Behrens, 2008).

I) DISPOSIÇÃO

Quando exploramos informação é muitas vezes necessário mexer na ordem em que ela nos é apresentada, através de acções directas - trazer ao de cima diferentes pormenores, roda-los, ver as coisas sob diferentes perspectivas. Através de novos pontos de vista sobre o mesmo conjunto de dados, podem-se revelar novas associações e alargar a compreensão destes dados através de uma visão mais aprofundada (Behrens, 2008).

⁴⁰ Retirada de: <http://www.directeffects.net/resources.html>

O Arranjo selectivo (*Selective Arrangement*) é uma opção simples e útil em estruturas que têm algum tipo de hierarquia, que à primeira vista pode não estar claramente discernida. Ao permitir o utilizador mudar automaticamente a organização dos elementos, estes podem facilmente passar de uma estrutura totalmente desorganizada e complexa, para uma hierarquia com níveis bem definidos, de onde transparecem novos pontos de destaque (Behrens, 2008).

As Colunas ordenáveis (*Sortable Collums*) são uma das ferramentas primordiais na análise de dados, e permitem ao utilizador escolher a ordem de agrupamento dos dados, consoante as variáveis de cada coluna da tabela (Behrens, 2008).

Quando se lida com informação multidimensional representada em sistemas de coordenadas bidimensionais, é impossível representar todas as dimensões dos dados, a não ser que o utilizador possa definir ele mesmo quais as dimensões a incluir nos eixos de coordenadas; a uma função de Dimensões Personalizáveis (*Custom Dimensions*), permite ao utilizador comparar, passo-a-passo, os vários pares de dimensões de dados (Behrens, 2008).

Por vezes, no decurso da análise de dados torna-se necessário pegar em dois elementos e comparar directamente os seus valores quantitativos de um modo isolado. Esta Comparação Isolada (*Isolated Comparison*), deve ser conseguida através de uma forma directa e sem ruído visual a distrair o utilizador dos dados que interessam (Behrens, 2008).

m) EXPLORAÇÃO

Aqui aborda-se o grau de detalhe a que o utilizador pode ter acesso, quando e onde quiser, em que medida é que este quer descobrir mais informação sobre determinado ponto. Esta questão é particularmente difícil porque envolve por um lado transmitir ao utilizador a sensação de liberdade autónoma para procurar o que quiser; por outro lado, transferir-lhe mais carga de trabalho na análise de dados. Este grau de liberdade deve à partida estar relacionado com o grau de familiaridade que o utilizador tem em lidar com o tipo de dados que implicam, e com a experiência que tem no uso da ferramenta de Visualização (Behrens, 2008).

Na maioria das vezes, é impossível mostrar todos os dados específicos a cada elemento da visualização de modo simultâneo, e é preferível adoptar o conceito dos Detalhes à medida (*Details on demand*), seguindo a metodologia de informação de *Shneiderman* (explicado na página 22). Esta estratégia de organização operacionaliza-se eficazmente quando os detalhes mais avançados relativos a uma área concreta, só aparecem quando solicitados. As Dicas (*Datatips*) são um exemplo simples desta medida, sendo activadas apenas quando o cursor paira sobre determinado elemento; surgem vulgarmente através de janelas de *pop-ups* e mostram informação extra relativa a esse elemento, permitindo ao utilizador explorar apenas dados que lhe interessem, quase sobre o mesmo princípio das camadas, mas específicas a elementos ou grupos concretos de informação (Behrens, 2008).

n) TRANSIÇÃO

A transição representa os aspectos dinâmicos da aplicação, e do *feedback* que o utilizador recebe, para perceber que se deram alterações de estados no sistema; para que este consiga acompanhar as movimentações dos dados, assim como as suas transformações. Esta comunicação visual é mais bem sucedida se se utilizarem animações que representem estas transições de estados na visualização. Tal como no mundo real, onde estamos habituados à sucessão de eventos ao longo de um espaço de tempo, onde nada muda de forma instantaneamente, é natural que o utilizador perca facilmente noção das transformações que se dão de modo imediato no sistema, sem algum tipo de pista perceptual que o alerte de que algo aconteceu (Behrens, 2008).

Interessa sobretudo que o utilizador perceba, através da animação, o que é que mudou na Visualização em relação ao que estava, e se possível como mudou. Tendo isto em conta há que ter a preocupação de criar animações que não sejam lentas ao ponto de quebrarem o ritmo da realização das tarefas, nem demasiado rápidas para serem percebidas pela mente humana; as transições mais simples, tal como a mudança de cor e tamanho, podem ser preferíveis às animações mais complexas e detalhadas.

2.5.3 PADRÕES DE INTERACÇÃO

Por fim, estes padrões representam os componentes mais funcionais da interacção do utilizador com a máquina. Estamos a falar das ferramentas de interface práticas por onde o utilizador vai comunicar os seus parâmetros ao sistema, operacionalizar o seu comportamento em acções de interacção na aplicação. Os padrões de interacção dividem-se nas seguintes categorias: Selecção Booleana, Ajuste Linear, Navegação espacial. Estas reflectem as diferentes formas do utilizador realizar tarefas de análise de dados (Behrens, 2008).

A Selecção Booleana segue o mesmo conceito da lógica algébrica, no sentido em que este tipo de selecção traduz-se facilmente em valores de verdadeiro ou falso. O utilizador selecciona de uma forma directa quais os parâmetros que quer activos ou inactivos na visualização, codificados através de variáveis dentro do conjunto de dados. Este tipo de selecção dá-se geralmente através de elementos da interface como os *radio buttons*, *checkboxes*, ou *dropdown menus* (Behrens, 2008).

O Ajuste Linear surge em contraste com a categoria anterior, orientada a dados que são estruturados de uma forma contínua e sequencial. Aqui as escolhas de sim e não podem ser demasiado precisas para lidar com estes dados – as escolhas devem ser mais analíticas. A ferramenta de Slide (*Slider*) é um elemento de interface que permite ao utilizador, através do arrasto do seu rato, navegar ao longo de toda uma série, parando no sítio correcto e aumentar o *zoom* de determinado elemento. Complementarmente, o Slide duplo permite elaborar ainda mais a selecção, pela definição de um intervalo, estabelecido pela escolha de um valor mínimo e valor máximo (Figura 18) (Behrens, 2008).

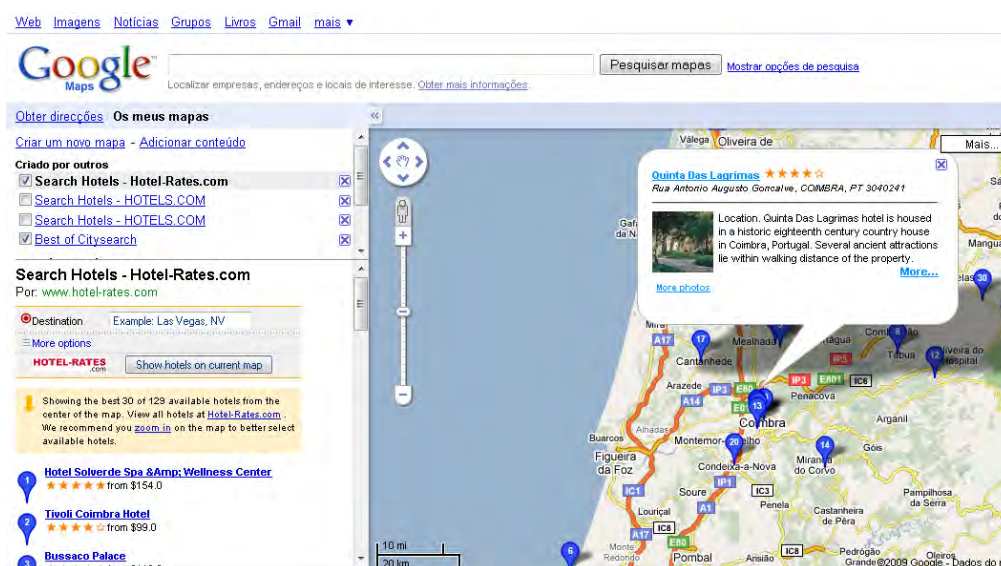


FIGURA 18 – A APLICAÇÃO GOOGLE MAPS⁴¹ REÚNE FERRAMENTAS COMO: DATATIPS, SLIDERS (DE ZOOM E DE TRANSLAÇÃO) E CHECKBOXES.

⁴¹ Disponível em: <http://maps.google.com/maps>

Por fim, a Navegação Espacial está intimamente ligada a Visualizações de configurações espaciais, que exibam comportamentos de navegação. Sucintamente, os mecanismos de navegação espacial estão presentes em qualquer Visualização que tenha alguns elementos espalhados “para fora” dos limites impostos pelo ecrã. O utilizador precisa de “deslocar-se” virtualmente de forma a encontrar e manipular estes elementos (Behrens, 2008).

A opção de Selecção e Arraste (*Drag and Drop*) é uma das mais conhecidas técnicas de navegação, seguindo metáforas do mundo real, onde pegamos e movemos objectos, tais como um mapa, de forma a alinharmos o que nos interessa ao nosso campo visual. Por vezes, a selecção e arraste carece de rigorosidade e precisão, servindo principalmente para deslocar-se a grande velocidade em distâncias maiores dentro da visualização (Behrens, 2008).

Esta opção está por exemplo disponível no portal *Google Maps*, utilizada para o utilizador se movimentar no mapa, para além de utilizar as setas direccionais (Figura 18, página 35).

Complementarmente, usa-se a Máscara de Selecção (*Selection Mask*) que permite ao utilizador, na mesma através do clique e arraste do rato, definir uma área de selecção, agrupando itens que lhe pareçam relacionados. Este acto de selecção de área pode resultar em acontecimentos automáticos tais como o *zoom* da selecção e detalhes à medida (Behrens, 2008).

2.6 USABILIDADE

A Usabilidade é definida pela ISO⁴² como: “a medida em que um produto pode ser usado por utilizadores específicos, de forma a alcançar objectivos com eficácia, eficiência e satisfação, num determinado contexto de uso” (ISO, 1998, p. 2).

Um dos problemas presentes da Visualização de Informação consiste em encontrar uma metodologia comum que permita de forma objectiva avaliar a Usabilidade das suas ferramentas. Esta dificuldade é facilmente percebida tendo em conta que estas ferramentas são normalmente desenhadas especificamente para cada caso, com dados, análises e mecanismos diferentes.

A rápida proliferação dos projectos de Visualização de Informação não foi acompanhada de um crescimento equitativo de estudos empíricos nesta área. Os problemas de usabilidade são ainda frequentemente encarados *a posteriori* no processo de implementação destes sistemas.

A falta de um critério standard para a avaliação de usabilidade resulta na criação de muitos projectos que, apesar de se auto designarem como Visualizações de Informação, falham em muitos dos seus propósitos da representação e análise de dados.

A grande maioria dos estudos pré-existentis sobre Usabilidade não está adaptada ao contexto da Visualização de Informação, sendo inadequados à morfologia específica dos seus sistemas; também o processo de interacção exploratória que se dá entre o utilizador e estes sistemas é difícil de avaliar através de métricas convencionais; este processo envolve mecanismos de percepção cognitiva complexos, relativos à forma como o utilizador compreende o que está a ver. Apesar de já existirem alguns trabalhos, tais como o de *Colin Ware* (2004) com uma extensa investigação sobre a percepção visual na Visualização de Informação, escasseia ainda a recolha de dados empíricos sobre os sistemas de visualização da nova geração; esta recolha deve incidir em tarefas cognitivas de maior complexidade envolvidas nestes sistemas - tais como a identificação de padrões emergentes através da leitura de séries temporais e o reconhecimento de grupos através da sua proximidade espacial (Chen, 2005).

⁴² ISO – International Organization for Standardization (ver mais em: <http://www.iso.org/iso/home.htm>)

2.6.1 USABILIDADE UNIVERSAL

Existe a compreensão generalizada entre os investigadores de que não existe nenhuma solução geral que sirva a todos os problemas de Visualização de Informação. Duas competências importantes ainda por atingir necessitam de ser estudadas a fundo antes de se poderem criar regras generalizadoras: a representação de grandes quantidades de informação e a capacidade de suporte para um público mais abrangente de utilizadores; este último é objecto de grande debate: ao mesmo tempo que os investigadores são encorajados a encontrar ferramentas mais abrangentes, que possam ser aplicadas a uma grande variedade de áreas, a falta de personalização resultante dessa generalização pode resultar em ferramentas inflexíveis pouco adequadas às necessidades particulares do utilizador (Plaisant, 2004).

A Usabilidade universal só pode ser vista a longo prazo, existem alguns tópicos de usabilidade que podem já ser adoptados para melhorarem as ferramentas de Visualização.

O primeiro passo lógico será melhorar o Design de interface de utilizador destas ferramentas. Existem já muitas teorizações desta área que podem ser úteis.

Um bom exemplo de usabilidade da interface do utilizador é a já referida aplicação *Map of the Market*⁴³, que através do aperfeiçoamento do Mapa de Árvore (Figura 19), conseguiu atingir o público geral. A hierarquia dos seus elementos é simples, mantendo o mesmo nível de profundidade, e a codificação das cores é intuitiva, utilizando a dicotomia conceptual vermelho-verde, compreendida pelos utilizadores. Também se tem o cuidado de manter um bom rácio da área dos rectângulos e o valor percentual que representam, de forma a não transmitir informação errónea que facilite a formulação de raciocínios incorrectos.

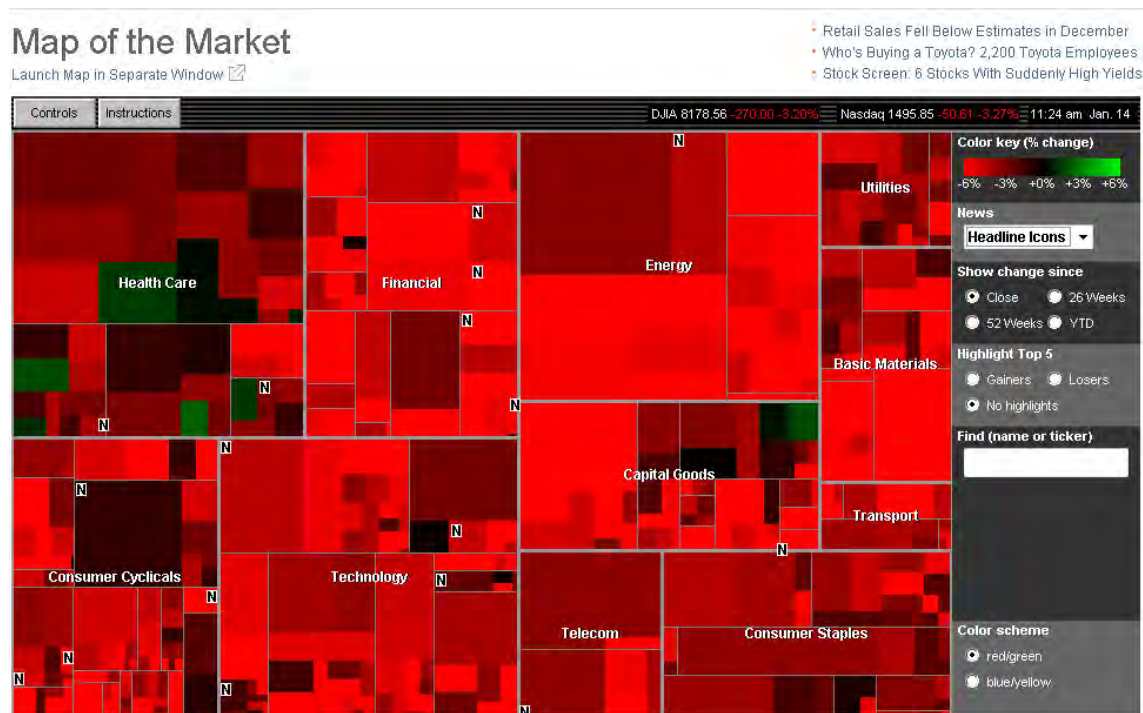


FIGURA 19 – APLICAÇÃO MAP OF THE MARKET⁴⁴, DO PORTAL ONLINE SMARTMONEY, É O EXEMPLO DE UM MAPA DE ÁRVORE COM SUCESSO.

O tipo de ajudas a fornecer ao utilizador é um aspecto importante, onde o maior desafio consiste em orientar pessoas normais para utilizarem estas ferramentas correctamente; tal facto deve ser tido em conta

⁴³ Disponível em: <http://www.smartmoney.com/>

⁴⁴ Retirado de: <http://www.smartmoney.com/map-of-the-market/>

quando se desenharem os mecanismos de interacção, seguindo o Mantra da Procura de Informação, discutido anteriormente.

É importante permitir ao utilizador a especificação personalizada dos seus parâmetros de pesquisa, direccionando assim as suas ferramentas. Uma boa ideia será possibilitar a escolha do tipo de correlações ou variações do padrão a procurar, assim como guardar um histórico de todo o processo de investigação, podendo voltar atrás sempre que se quiser; para além disso o histórico permitirá ao utilizador criar métodos automatizados baseados no que fez anteriormente, podendo utiliza-los em novos conjuntos de dados.

Segundo *Stephen Few*, nós usamos as representações visuais com dois objectivos principais: a procura de padrões significativos e da sua análise extensa quando eles são encontrados, e a comunicação do conhecimento apreendido a outras pessoas (Few, 2006). Outro ponto relevante reside então na integração de ferramentas de comunicação social, permitindo a partilha de dados e visualizações assim como a sua construção conjunta com outros investigadores e colaboradores.

Tal como em qualquer aplicação multimédia, devem-se considerar os requisitos de *hardware* necessários à boa utilização da aplicação. Considerando o público-alvo, devem-se estimar quais os tipos de computadores de onde vai ser acedida a aplicação, assim como as velocidades de ligação, e tomar as medidas necessárias para que não se dêem grandes períodos de espera. O desafio da usabilidade universal ainda é mais complicado se pensarmos nos utilizadores com ligações de rede lentas, ecrãs pequenos, ou em aplicações adequadas a dispositivos móveis, sendo este último um campo quase inexplorado na Visualização de Informação.

A Visualização de Informação está cada vez mais acessível aos olhos do público geral, sendo cada vez mais crucial o estudo da usabilidade universal neste contexto. O que está feito ainda está longe de suficiente. Ainda mais ao abandono se encontram as questões de acessibilidade, visualizações que tenham em conta utilizadores com necessidades especiais; como é que se representam dados a utilizadores daltónicos, ou até mesmo utilizadores invisuais? São perguntas que devem ser ainda mais exploradas – e é através de aplicações que primam por um alto rigor de acessibilidade, que provavelmente surgirão grandes conceitos de usabilidade da Visualização de Informação.

Seguindo o que foi discutido anteriormente, através dos padrões de design de *Behrens*, ou do *Mantra da Procura de Informação* de *Shneiderman*, a usabilidade em sistemas de Visualização de Informação implica à partida a distinção temática entre dois campos principais, envolvidos em qualquer das suas ferramentas: a representação visual e os mecanismos de interacção, que se regem sob conceitos distintos e como tal têm de ter diferentes medidas de avaliação.

2.6.2 ESTÉTICA DA INFORMAÇÃO

Dentro da área da representação visual, distingue-se a estética da informação como tendo um destaque especial, tendo sido objecto em vários estudos, incidindo principalmente nas redes de grafos. Existem vários estudos que demonstram que uma boa qualidade estética reduz o tempo e número de erros na execução de tarefas, melhorando a eficiência e eficácia (Stasko, Catrambone, Guzdial & McDonald, 2000) (Ngo & Byrne, 2001).

Um estudo mais recente (2007) efectuado por *Nick Cawthon* e *Andrew Vande Moere* procura estudar as discrepâncias de eficácia e eficiência existentes entre várias tipologias de estruturas em árvore; constatou-se que as medidas de eficácia das estruturas tridimensionais eram consideravelmente inferiores às estruturas bidimensionais. A aplicação *SunBurst* (Figura 20) foi a que registou valores maiores de eficácia e eficiência, ao mesmo tempo que foi considerada pelos utilizadores como a mais bonita; a sua percentagem de tarefas efectuadas correctamente superou a do *Windows Explorer*, ultrapassando a familiaridade já existente deste sistema.

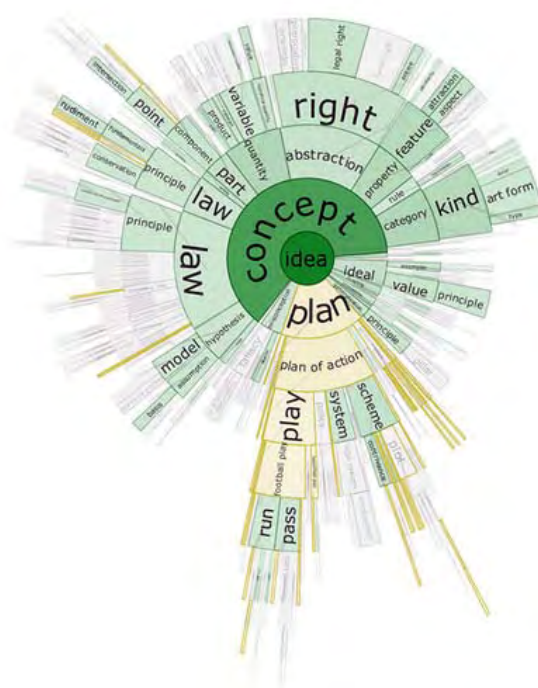


FIGURA 20 – EXEMPLO DA TIPOLOGIA *SUNBURST*⁴⁵, DO PROGRAMA *DOCUBURST* (COLLINS, 2006), UMA VISUALIZAÇÃO DE DOCUMENTOS.

Existiram no entanto sistemas que apesar de deterem grandes valores relativos à sua beleza percebida pelos utilizadores tiveram fracas prestações na realização de tarefas. Para justificar estes resultados, este estudo apoia-se na teoria de que a eficácia e eficiência de determinadas técnicas de visualização variam consoante os tipos de dados com que trabalham; para além disso o factor da interacção é também decisivo para a estética percebida (Cawthon & Moere, 2007). Quanto mais bonita for considerada a visualização, maior será o grau de paciência que o utilizador terá para aprender a utilizar a ferramenta; cria-se uma relação emotiva com o objecto, que estimula o pensamento criativo para resolver soluções, associando-se às noções de Design Emocional (Norman, 2004).

Existe sobretudo a necessidade de se desenharem aplicações com o intuito de as tornar no mais transparente possível para o utilizador; em vez de funcionar como uma fronteira que separa o utilizador da informação, o elemento mediador entre a informação e o utilizador deve ser quase imperceptível, dando a sensação de que não está a usar ferramenta nenhuma, de que é transparente (Few, 2006).

2.6.3 TÉCNICAS DE AVALIAÇÃO

A investigadora *Carla Freitas et al.*, apresenta uma proposta para um modelo de técnicas de avaliação de usabilidade, que foi posteriormente utilizado para estruturar a avaliação da usabilidade de uma aplicação de Visualização (Freitas, Luzzardi, Cava, Winckler, Pimenta & Nede, 2002).

Este modelo é interessante para esta investigação porque consegue estruturar, de forma objectiva, um conjunto de orientações para a avaliação de usabilidade; estas orientações derivam de propósitos comuns já referidos, dividindo as ferramentas de Visualização de Informação em dois conjuntos de critérios: representações visuais, e de mecanismos de interacção.

⁴⁵ Retirado de: <http://www.flickr.com/photos/seeminglee/2218565333/>

a) REPRESENTAÇÕES VISUAIS

A representação visual divide-se em várias subcategorias distintas: limitações, complexidade cognitiva, organização espacial, codificação da informação, transição de estados:

As limitações dizem respeito a restrições visuais tais como o tamanho do ecrã e o número máximo de elementos passíveis de serem mostrados.

A complexidade cognitiva da imagem diz respeito à caracterização dos dados e à maneira como estes são destacados, medidos através da densidade e dimensão dos dados e as formas de destacar informação relevante.

A organização espacial está ligada à forma geral da interface da aplicação, e no grau de facilidade com que se encontram os elementos, tanto através da sua posição com da sua orientação espacial. Devem-se ter em conta aspectos como a oclusão de objectos sobrepostos assim como formas de ordena-los logicamente, assim como se enquadra determinado objecto em destaque através das referências contextuais à sua volta.

O conceito de codificação da informação está ligado com o mapeamento dos dados em elementos visuais, e com os símbolos adicionais relacionados a aspectos da realidade, e a maneira como estes são utilizados para representar a informação.

Finalmente, a transição de estados consiste na análise do que acontece na aplicação quando o utilizador muda dados de entrada (mudança de *zoom*, mudança de camadas de visualização). O tempo que leva para se tornarem visíveis essas mudanças é um factor de medida, assim como se há consistência na orientação visual dos elementos, antes e depois da transformação.

b) MECANISMOS DE INTERACÇÃO

Este bloco conceptual diz respeito às funcionalidades que permitem ao utilizador realizar tarefas de interacção e manipulação dos dados. Estes mecanismos separam-se em três categorias distintas: Orientação e Ajuda, Navegação e Consulta, Redução de dados:

A Orientação está subjacente no conjunto de controlos existentes que permitem ao utilizador escolher o nível dos detalhes a serem mostrados na visualização. A ajuda implica o estudo das ferramentas de apoio que são disponibilizadas, tanto para a recapitulação ou correcção de acções tomadas (opção de *undo* e *redo*) como para a mostragem de informação adicional que possa apoiar as decisões do utilizador.

A navegação diz respeito aos mecanismos existentes para que o utilizador desloque o seu ponto de vista na visualização, assim como a selecção de objectos ou conjuntos de objectos específicos. Tal poderá implicar a manipulação geométrica dos elementos, ou a mudança de pontos de vista da visualização e o aumento ou diminuição dos elementos (*zoom*). A Consulta está mais ligada à capacidade de procura directa sobre perguntas concretas postas pelo utilizador, através da manipulação directa dos parâmetros de entrada, ou através da introdução de palavras-chave.

A Redução de dados está ligada aos processos de filtragem, que diminuem a amostra de dados tendo em conta as necessidades de visualização focalizada só em alguns aspectos. Também nela estão inseridos tarefas como o agrupamento de dados, que permite a segmentação da amostra de dados através de símbolos, assim como o refinamento dos dados, através de cortes simples de informação irrelevante.

3. INVESTIGAÇÃO EMPÍRICA DO ESTUDO DE CASO

Após a introdução dos conceitos teóricos, será relevante apresentar de que forma é que este conhecimento foi conjugado com o estudo de caso da investigação.

3.1 APRESENTAÇÃO DA METODOLOGIA DE INVESTIGAÇÃO

Para definir mais especificamente a tipologia desta investigação podemos utilizar dois critérios de classificação: propósito e método. (Carmo & Ferreira, 1998)

Segundo o autor, a classificação quanto ao propósito tem sobretudo em conta a aplicação posterior que os resultados da investigação poderão ter, assim como o grau de generalização que estes possam ter para a população do estudo.

Neste ponto de vista, são identificadas duas classificações passíveis de serem atribuídas a este trabalho: Investigação Aplicada e projecto de Investigação e Desenvolvimento (*I&D*).

A definição de investigação aplicada dita que esta deve acabar pelo teste de uma teoria, aplicando-a um contexto real. Por outro lado, um projecto de *I&D* implica o desenvolvimento de uma aplicação com o intuito de construir um produto a ser utilizado para determinados fins.

Apesar de ter algumas semelhanças com um projecto de *I&D*, a presente investigação não se enquadra como tal, o seu produto final não tem como objectivo ser um produto comercial acabado, mas antes um ponto de partida para futuros desenvolvimentos. Considera-se então mais apropriado classificar a presente investigação como investigação aplicada, na medida em que esta começa com a formulação de uma estratégia de uma visualização, que é então inserida no contexto do portal Sapo.

Quanto ao método, tem-se em consideração a estratégia utilizada no decorrer da investigação, que está dependente dos tipos de questões que esta pretende responder.

Aqui enquadrámos esta investigação na categoria de estudo de caso. Para compreender esta classificação deve-se entender o estudo de caso como um método científico que estuda um determinado fenómeno dentro do seu contexto (Yin, 1988), em oposição a outras investigações que separam o fenómeno em estudo do seu contexto específico. Para esta investigação entende-se o fenómeno como o *Broker Sapo* e a sua utilização no contexto da Edição do Portal Sapo.

Yin (1988) explica ainda que é comum para este tipo de investigações a utilização de diferentes fontes de dados, onde os dados recolhidos podem ser de natureza qualitativa, quantitativa ou ambas. (Yin, 1988)

Na utilização de diferentes fontes de dados, esta investigação faz uma análise documental de conteúdos abrangidos pela área da Visualização de Informação; explorando as diferentes publicações existentes na matéria, enquanto procura identificar quais as teorias mais importantes e relevantes para a sua componente prática (protótipo funcional); tendo isto em conta pode-se também definir esta investigação como exploratória (Loureiro, 2006). Enquanto que a pesquisa bibliográfica tem certamente um papel decisivo no seu enquadramento teórico, o estudo de campo é também importante, a ser utilizado no decorrer da avaliação do protótipo posteriormente implementado.

Quanto à obtenção e ao tratamento de dados (Pardal & Correia, 1995), a investigação enverga a natureza quantitativa e qualitativa, porque tanto vai implicar a análise estatística de variáveis referentes à avaliação da usabilidade do protótipo, assim como a análise de conteúdo relativa à opinião do público-alvo – a redacção/edição do portal Sapo.

3.2 PLANO DE ACÇÃO

A investigação descrita neste documento procura chegar à criação de uma aplicação de visualização e análise de dados, que transforme registos de pesquisas efectuadas num portal *web* em informação de fácil leitura, que auxilie a tomada de decisões. Como ponto de início, esta investigação centra-se num público-alvo específico, a equipa de Edição do Portal Sapo, e esta finalidade implica que a aplicação vá utilizar dados dentro deste contexto – neste caso do sistema *Broker Sapo*. Esta centralização do público-alvo permite a construção de uma aplicação que esteja orientada para determinadas necessidades do utilizador. Por outro lado, devido às potenciais áreas de aplicação deste protótipo, este poderá eventualmente adaptado a outros contextos de uso, devendo ser construído com o objectivo de não ser um produto acabado, mas um ponto de partida.

Esta investigação (assim como qualquer outra) implica a elaboração de um plano de acção, que a oriente de maneira a criar uma sequência lógica, onde a fase anterior influencia a fase seguinte.

A primeira fase traduz-se na definição da pergunta de investigação, assim como das finalidades e objectivos da investigação, em que o único conhecimento pré-existente é acerca do sistema *Broker Sapo* e do seu potencial. Para a realização desta fase foi necessária uma entrevista exploratória, que se deu com o Editor-chefe do Portal Sapo. Esta entrevista serviu para que se fossem indagadas as eventuais necessidades da equipa da redacção, tendo em conta o seu dia-a-dia profissional e a sua actual relação com a informação do *Broker*. Por outro lado, esta entrevista também concedeu aos investigadores a oportunidade de analisarem os sistemas de visualização já pré-existent – quais as suas vantagens e limitações. Com base na informação recolhida foi possível formular: refinar a pergunta de investigação, definir as finalidades e esboçar os objectivos da aplicação.

A segunda fase constitui a componente teórica da investigação, fulcral para a solidificação da sua parte prática: a revisão da literatura, respeitante à área de estudo. Esta fase tem como objectivo agregar um

conjunto de informação que possa ser valiosa para a conceptualização do protótipo, assim como para a interpretação dos dados (Merriam, 1988). Passa pela introdução teórica da área de investigação, das suas particularidades, e de pontos específicos ao contexto desta investigação. A conclusão desta fase ajudou à redefinição dos pontos definidos na fase anterior, principalmente dos objectivos da aplicação.

A terceira fase envolve duas actividades principais, que não estando directamente relacionadas, os seus respectivos resultados convergem na próxima fase. Estas actividades são: a conceptualização das metáforas visuais e a formação auto-administrada do uso da ferramenta *Processing*, a ser utilizada posteriormente na implementação do protótipo. As duas aparecem na mesma fase da investigação porque são realizadas simultaneamente; enquanto se vão realizando experiências técnicas, também vão sendo realizadas experiências conceptuais. No final desta fase, os investigadores são capazes de criar uma aplicação avançada com a ferramenta técnica, assim como já terem definido um esquema visual para o protótipo.

A quarta fase corresponde à implementação técnica da aplicação, utilizando elementos visuais da fase anterior. São dados vários graus de importância a diferentes componentes passíveis de serem programadas, sendo as mais importantes as primeiras a serem implementadas. De um modo geral, começou-se com a recolha e tratamento dos dados, seguindo-se com a visualização de uma amostra de dados, e por fim a implementação de ferramentas de interacção.

Por fim, após a conclusão de um protótipo funcional, é seguida a fase de avaliação, onde o protótipo e o público-alvo (a equipa de edição do Portal Sapo) se reúnem, e dessa interacção são registados os problemas encontrados e as sugestões oferecidas. Formula-se uma ferramenta de recolha de dados, e efectua-se a recolha e análise dos dados. Com base nesta análise são desenvolvidas as conclusões e sugestões futuras.

3.3 CARACTERIZAÇÃO DO PÚBLICO-ALVO

O público-alvo para esta investigação pode-se identificar, num ponto de vista mais abrangente e generalizador, como qualquer equipa responsável pela dinamização e actualização de um produto/serviço/actividade, onde os dados de pesquisas *Web* possam ser úteis para a tomada de decisões, através da detecção e análise de tendências dos utilizadores.

Num sentido mais restringido à aplicação concreta do protótipo que resulta de toda a investigação, o público-alvo define-se como os elementos da equipa de edição do Portal Sapo, que possam ter acesso à informação das pesquisas *Web* realizadas no Portal, disponibilizada pelo sistema *Broker* da mesma empresa. Dentro deste público-alvo contam-se jornalistas, editores, e técnicos de publicidade. Presume-se que este público-alvo já tenha um nível elevado de competências técnicas, pelo menos na utilização de aplicações multimédia, *tanto offline e como online*.

A amostra dos participantes na avaliação do protótipo é composta por 7 elementos da equipa de edição do Portal Sapo, com idades compreendidas entre os 22 e os 52 anos, 2 membros do sexo masculino e 5 membros do sexo feminino. Relativamente à sua função profissional, contam-se 4 Jornalistas, 2 Gestores Editoriais e 1 Editor.

Para a escolha dos participantes, utilizou-se uma amostragem não-probabilística, e por conveniência. (Carmo & Ferreira, 1998) Isto implica que estes tenham sido seleccionados com base sobretudo na sua disponibilidade, mas também na secção do Portal onde trabalham. Esta selecção teve a co-orientação do Editor-chefe do portal, que indicou quais os participantes que poderiam ser mais relevantes de inquirir.

3.4 CONTEXTUALIZAÇÃO TÉCNICA

É relevante clarificar alguns dos conceitos técnicos que são amplamente referidos nesta investigação, que merecem especial destaque, nomeadamente, o *Processing* e o *Sapo Broker*.

O *Processing*⁴⁶ é um projecto *open source*, inicialmente criado em 2001 por Ben Fry e Casey Reas, com o propósito inicial de ser um *software* educativo, usado como ferramenta de apoio para o ensino dos fundamentos de programação num ambiente visual.

Actualmente é um ambiente de desenvolvimento com uma linguagem própria de programação (baseada em Java), utilizado já para o desenvolvimento de produtos comerciais, especialmente na área da Visualização de Informação, das instalações educativas, e arte digital.

É executável em ambiente *Windows*, *Mac* e *Linux*, e permite a exportação dos projectos como aplicações *online*, ou para aplicações *offline*, compatíveis com os sistemas referidos anteriormente. É suportado por uma larga comunidade *online*, onde são disponibilizadas várias extensões e códigos livres.

A escolha para a sua utilização foi tomada tendo em conta a sua versatilidade e quantidade de aplicações de Visualização de Informação construídas neste ambiente.

O *Sapo Broker*⁴⁷ é uma plataforma de distribuição de mensagens, um programa de *Message Broker*, que permite a subscrição de vários tópicos (através do seu componente público), representativos de dados relativos, em tempo real ao portal Sapo. É escrito em Java, mas dispõe de várias bibliotecas de clientes em várias linguagens (*Perl*, *Python*, *PHP*, *Ruby*, *.Net*, *C* e *Erlang*).

3.5 DESIGN FUNCIONAL

A tomada de decisão sobre qual a metáfora ou metáforas visuais a utilizar é usualmente um processo iterativo e experimental. Apesar do enquadramento teórico exaustivo que procura auxiliar a objectivação esta decisão, existe sempre uma componente subjectiva associada.

Os agentes concretos mais relevantes que podem servir como ponto de partida para esta etapa são os tipos de dados que se pretendem visualizar, e as suas características ajudam a definir quais as visualizações que podem ser mais adequadas.

Podemos assim identificar a relação directa entre o seguinte par específico de variáveis: as palavras-pesquisa e o número de pesquisas. Esta correlação já está implícita no sistema de visualização de pesquisas pré-existente, onde nos é apresentada uma listagem dos termos mais pesquisados durante a última janela temporal. Conjuntos de dados deste género são facilmente transcritos sob a forma de tabelas, e é natural imaginar a construção de uma matriz de dispersão, ou até num gráfico de bolhas (tal como no exemplo da Figura 9, na página 24) - se quisermos incluir mais do que duas variáveis na correlação – associados a estes dados.

Complementarmente, a amostragem de dados é discreta, onde não existe um espectro de resultados registados continuamente, mas através de intervalos temporais. Por cada janela temporal regista-se a soma do número de pesquisas por palavra.

Existe ainda outro tipo de dados que se pode incluir neste projecto, tendo em conta os objectivos iniciais propostos: as configurações espaciais. Tendo em conta a existência das referências geográficas associadas às pesquisas, é natural imaginarmos alguma forma de mapa que traduza esta informação.

⁴⁶ Disponível em: <http://processing.org/>

⁴⁷ Disponível em: <http://softwarelivre.Sapo.pt/broker>

Considerando estas diferenças de dados relevantes para o projecto em desenvolvimento, o processo de selecção de uma metáfora visual dividiu-se em dois objectos de estudo principais. O primeiro objecto de estudo diz respeito à representação da correlação pesquisas versus número de ocorrências, e de como traduzir a alteração dos estados dentro de uma janela temporal. O segundo objecto de estudo focaliza-se na referência geográfica que estas pesquisas possam ter, procurando criar um mapa nacional que reflecta a distribuição do total de pesquisas efectuadas por região, mas também que reflecta a distribuição nacional para cada palavra pesquisada.

3.5.1 PESQUISAS

Inicialmente foram realizadas algumas experiências para a representação de palavras-chave, adoptando alguma da morfologia do visualizador de pesquisas pré-existente do *Broker Sapo*.

Uma destas experiências consiste na listagem das cidades do país, alinhadas pelo número de pesquisas nelas registadas (Figura 21, página 45). O tamanho da fonte de cada linha desta lista é directamente proporcional ao número das pesquisas que ilustra. A lista é dinâmica no espaço, movimenta-se verticalmente como uma cascata, onde o movimento das suas linhas não é uniforme. A velocidade de cada linha está relacionada principalmente com o número das pesquisas que representa, sendo ainda acrescentado um factor aleatório na velocidade, que tem um peso menos significativo na movimentação.



FIGURA 21 - PRIMEIRA EXPERIÊNCIA NO PROCESSING: LISTAGEM DE PALAVRAS COM OS RESPECTIVOS NÚMEROS DE OCORRÊNCIAS.

Neste *layout* observa-se que as cidades com menos pesquisas, movimentam-se com maior velocidade, sendo menos legíveis e pouco relevantes na vista geral da visualização. Por outro lado, as cidades com mais pesquisas têm menor velocidade, o que resulta que estas tenham destaque maior.

No entanto, esta visualização não trouxe grande novidade no ramo, na medida em que não traz grandes vantagens em relação a uma lista comum. É no entanto um bom ponto de partida, tanto a nível conceptual como a nível técnico. A nível conceptual porque introduz a ideia de movimento como factor de distinção e a nível técnico é no fundo um bom exercício para se ambientar com a ferramenta (*Processing*).

Após esta exploração preliminar, começou-se a ponderar que uma solução baseada em *tag-clouds*, ou listagens de palavras, o que se revelou uma opção pouco atraente, pois limitava um pouco as opções de representação; ao termos uma visualização cujo maior relevo está nas palavras e não em ícones visuais, ficamos muito dependentes do tamanho das palavras, cuja combinação pode resultar numa visualização

demasiado heterogénea; também existe a especial dificuldade em sobrepor várias palavras sem perder legibilidade e simplicidade.

Procurou-se sobretudo orientar esta investigação num caminho visual mais directo e simplista, cujo carácter interactivo pudesse complementar o menor grau de detalhe subjacente.

Um dos projectos pré-existentes que serviu de inspiração inicial é uma visualização dos objectivos traçados pelos utilizadores do portal *43Things*⁴⁸, onde estes enumeram 43 tarefas que querem realizar (Figura 22).

O conceito por detrás da visualização em si é simples: o borbulhar de bolhas representativas destas tarefas mais recentes, onde o tamanho da bolha é representativo do número de utilizadores que as pretendem realizar.

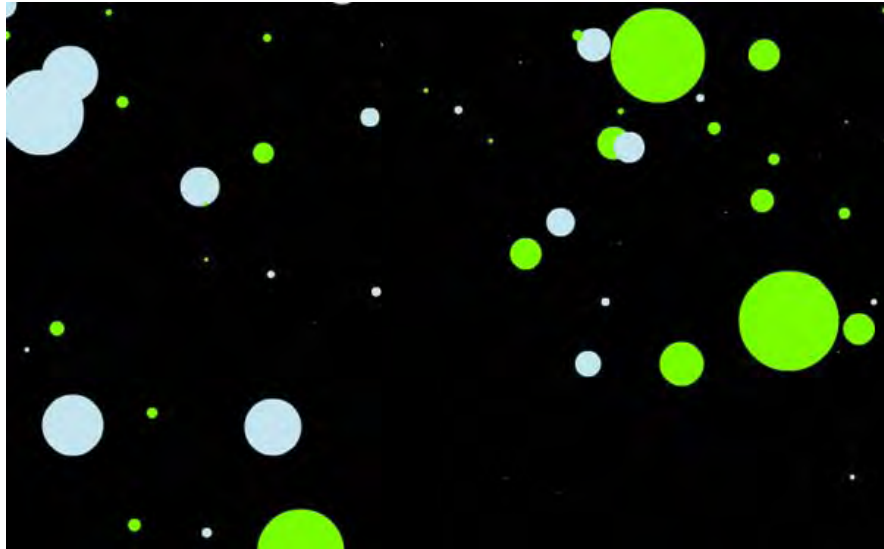


FIGURA 22 – PROJECTO 43THINGS⁴⁹.

Não existe grande profundidade ou detalhe nos dados que representa, sendo apenas a dimensão das bolhas utilizada para os representar quantitativamente. A posição horizontal das bolhas é atribuída aleatoriamente, e a posição vertical é ditada ao longo do tempo por uma determinada velocidade, também atribuída aleatoriamente. No entanto, esta simplicidade traduz bem um ambiente fluído e consistente, tornando-se intuitiva a sua navegação - utiliza o ambiente das bolhas de ar e associa-lhe o contexto do *43Things*, tal como podia utilizar outro indicador. Esta associação intuitiva tornou-se então num bom ponto de partida para o contexto desta investigação.

Primeiro que tudo, procurou-se recriar o mesmo efeito alterando apenas o conjunto de dados. Esta primeira experiência serviu também para avaliar o grau de complexidade técnica que tal visualização poderia requerer do programador, e calcular a viabilidade de elaborar uma versão mais complexa, se necessário.

⁴⁸ Disponível em: <http://www.43things.com/>

⁴⁹ Disponível em: <http://projects.flowingdata.com/43things/>

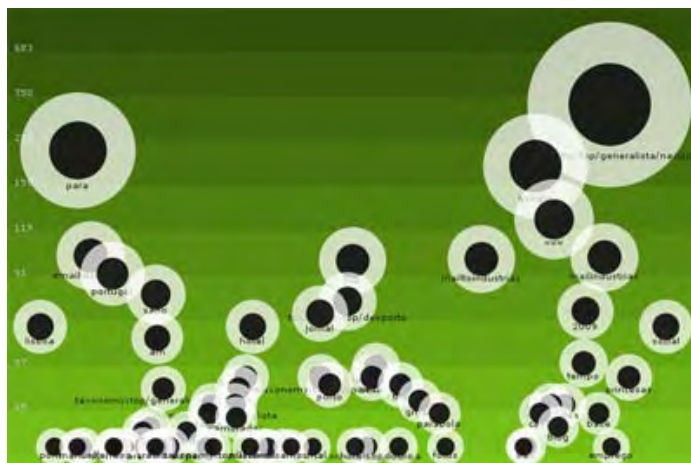


FIGURA 24 – ESQUEMA VISUAL ADOPTADO PARA A VISUALIZAÇÃO.

Para auxiliar a percepção das diferenças de altura entre os diferentes ovos foi adicionado um gradiente de cores. Este consiste em faixas rectangulares, de tamanho regular, cuja cor vai diminuindo de luminosidade consoante a altura da visualização. Pretende-se com este pano de fundo criar linhas orientadoras que sirvam como uma régua ao olho humano; ao mesmo tempo procura-se acentuar o destaque aos ovos em alturas superiores, cuja cor branca que os circunscreve, em contraposição com uma cor de fundo mais escura, cria a ilusão de maior destaque.

Uma legenda com os valores do número de pesquisas correspondentes a intervalos definidos de altura foi incorporada do lado esquerdo da visualização, à semelhança da legenda de um gráfico bidimensional.

3.5.2 MAPA GEOGRÁFICO

Tal como no processo de design funcional das pesquisas (página 45), começou-se por analisar o sistema pré-existente para lidar com os dados georreferenciados. Este sistema consiste num mapa do país dividido por distritos, onde cada distrito indica através de texto a percentagem de pesquisas efectuadas, em relação à amostra total, dentro do último intervalo temporal.

Procurou-se de imediato adicionar um elemento visual representativo dessas percentagem, de modo a criar uma mapa que permitisse de um modo breve perceber quais as maiores regiões de pesquisa.

Neste sentido surgiu uma primeira influência que é a visualização dos hábitos de consumo pelos países do mundo do *New York Times*⁵¹. Este conceito mais complexo resume-se a um mapa-mundo, cujos países são apenas representados por rectângulos, cujo tamanho altera conforme os valores de cada um relativo a hábitos de consumo. A posição destes rectângulos é a única verdadeira ligação ao país que representam, pois o delineamento geográfico é descartado completamente. Mesmo esta posição não é estática, variando ligeiramente com o intuito de manter a proximidade entre os vários rectângulos.

Apesar de ser apelativa esta desmaterialização completa do mapa geográfico real, mantendo a sua identidade geográfica apenas pela sua posição relativa, acabou por se considerar que esta opção teria uma leitura muito difícil para o caso de Portugal. Tal deve-se ao facto de existir uma enorme centralização de pesquisas em torno das regiões do Porto e principalmente Lisboa, tornando outras regiões, com um número mais reduzido de ocorrências, praticamente invisíveis.

No entanto, foi retirado deste exemplo o tamanho como elemento de diferenciação. É então considerada a ideia de dispor sobre o mapa do país elementos homogéneos que são maiores consoante o número absoluto de pesquisas que representam.

⁵¹ Disponível em: <http://www.nytimes.com/interactive/2008/09/04/business/20080907-metrics-graphic.html>

Tendo este elemento em conta, surgiu outra possível alternativa numa visualização que mede as taxas de pobreza nos Estados Unidos desde 1980 a 2007⁵². Aqui mantém-se o tamanho como indicador representativo, neste caso do raio dos círculos que pairam sobre cada estado do país. Desenvolveu-se então à sua semelhança uma visualização geográfica inicial, seguindo as mesmas especificações (Figura 25, à esquerda).

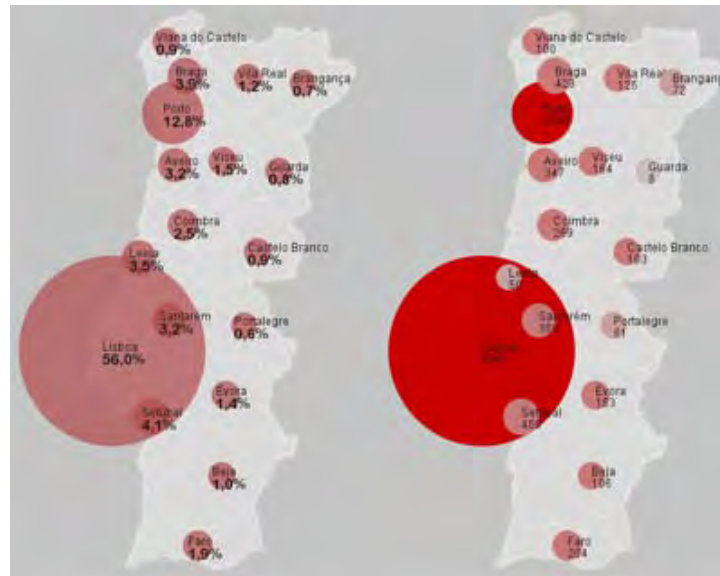


FIGURA 25 – UMA DAS PRIMEIRAS EXPERIÊNCIAS COM AS DISTRIBUIÇÕES GEOGRÁFICAS. UMA UTILIZA APENAS O TAMANHO COMO FACTOR DIFERENCIADOR E A OUTRA ADICIONA A MUDANÇA DA SATURAÇÃO DA COR.

Identificou-se nesta visualização o problema de diferenciar elementos que se sobrepunham espacialmente, tais como se pode verificar para os distritos de Santarém e Setúbal, que neste conjunto de dados são englobados totalmente por Lisboa. Tentou-se contornar este problema através da inserção de outros factores de diferenciação, tais como luminosidade ou saturação (Figura 25, à direita). O problema continuava a não ser bem resolvido, e procuraram-se então formas alternativas de o abordar.

Ponderou-se sobre uma solução que utilizasse o 3D para possibilitar a passagem do número de pesquisas para um eixo Z da visualização⁵³. Porém, esta hipótese foi posta de parte, na medida em que poderia ser demasiado complexa na sua implementação técnica. Por outro lado, foi considerada fraca a aplicabilidade do 3D para este projecto, na medida em que a introdução de um mapa 3D poderia complicar desnecessariamente a visualização de dados simples. Dentro do contexto de uso, o mapa a 3 dimensões não traria grandes benefícios em relação a uma solução a 2 dimensões.

Optou-se então por um sistema mais convencional, mas que mantém um grau de efectividade e clareza elevado. Consiste na coloração de cada distrito com a mesma tonalidade de cor, mas alterando a luminosidade consoante uma escala discreta, onde cada item da escala representa uma gama de valores definidos de pesquisas. Por outras palavras, os distritos com uma gama maior de pesquisas encontram-se com uma cor mais escura, e vice-versa. A cor foi considerada para se integrar no contexto institucional Sapo, optando-se pela tonalidade verde, semelhante à tonalidade usada na visualização das pesquisas (Figura 26).

Posteriormente, considerou-se mais relevante escalar os resultados por percentagem, em alternativa aos números de pesquisas. Com o clique directo sobre cada região seria então possível ver mais detalhes. A ideia inicial seria que este clique desdobrasse mais informação específica para esse distrito, nomeadamente número de pesquisas, o top de pesquisas mais efectuadas, e últimas datas de pesquisas.

⁵² Disponível em: <http://projects.flowingdata.com/poverty/>

⁵³ Um exemplo semelhante pode ser consultado em: http://www.time.com/time/covers/20061030/where_we_live/

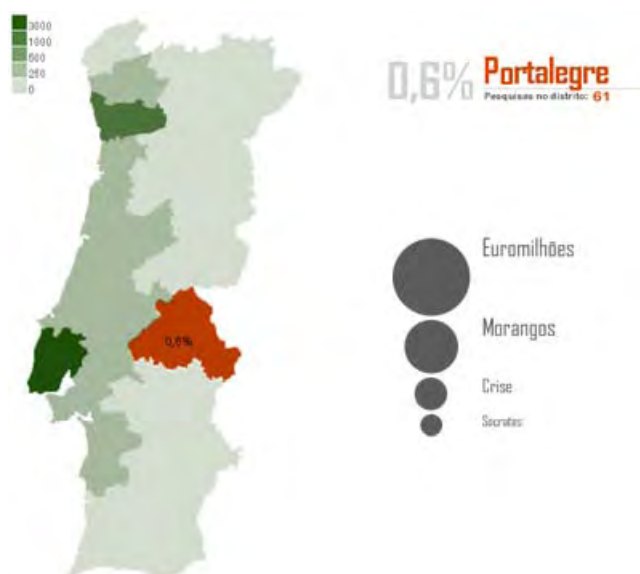


FIGURA 26 – UMA DAS TIPOLOGIAS A SEREM ESTUDADAS, QUE PERMITEM VER MAIS DETALHES (À DIREITA) RELATIVOS A CADA REGIÃO.

Apesar da implementação técnica desta solução ter sido bem sucedida, tornou-se a certa altura evidente que se tinha gerado um grande distanciamento entre a visualização das pesquisas (através dos ovos) e esta componente geográfica. Considerando que se pretendia a construção de um protótipo que incorporasse estas duas funcionalidades, era necessário conciliar estas duas visualizações de forma a não haver um choque visual entre os dois.

Decidiu-se então optar por utilizar esta distribuição geográfica associada a palavras-pesquisa, enquadrada na informação resultante do clique directo nos elementos, seguindo uma estratégia de detalhes à medida (*details-on-demand*). Ao clicar num dos ovos, representativo de uma palavra-pesquisa, é mostrado o mapa regional das pesquisas efectuadas sobre essa palavra (Figura 27).

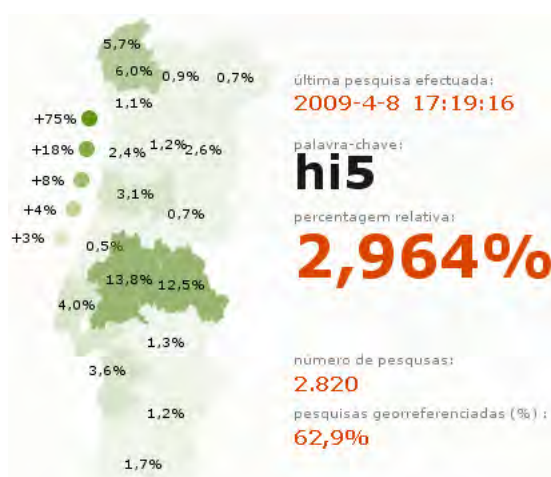


FIGURA 27 – CONTEÚDO FINAL DISPONIBILIZADO NOS DETALHES À MEDIDA DE CADA OVO.

A opção que incorpora um mapa sobre as pesquisas gerais, onde a informação de *pop-up* incluísse o top de pesquisas (Figura 26) talvez fosse a mais funcional, todavia acabava por ser a mais convencional. Preferiu-se assim enveredar por um caminho mais experimental, onde se desse maior ênfase às palavras pesquisadas e à sua relação entre si.

3.5.3 ESTUDO DE INTERFACE

Mesmo projectando este protótipo com o intuito de ser uma aplicação de análise de dados sobretudo visual, tornou-se importante decidir quanto ao seu possível carácter interactivo. Por um lado existia a opção de desenvolver um sistema cujo utilizador fosse passivo, usado para visualizar de uma forma directa e menos exaustiva a informação. Isto à semelhança do sistema anterior usado na edição do Portal Sapo, onde não existe qualquer interacção com o utilizador. Por outro lado, podia-se optar por uma solução interactiva, que implicasse exploração e tomadas de decisão do utilizador, de forma a chegar à informação pretendida.

Esta última foi a opção a ser adoptada, com base na informação que foi adquirida na entrevista exploratória com o Editor-chefe do Portal, que se mostrou mais interessado em aplicações interactivas.

O processo de conceptualização das ferramentas de interacção teve começo após a definição da metáfora visual da visualização. No entanto, já se encontrava presente desde o início da investigação que a ferramenta interactiva principal da visualização deveria ser o clique directo nos ovos, porque seria intuitiva, aproximando-se mais da ideia de manipulação directa dos dados, de ver e tocar no que se vê.

Foram realizadas poucas experiências sobre o design do feedback que esta interacção provocaria; o desdobramento de uma nova secção lateral na aplicação, assim como a abertura de janelas externas foram alguns exemplos. No entanto, faltava nestas hipóteses uma relação perceptível entre a informação de destaque e o ovo correspondente. Procurou-se então manter esta relação através de uma posição espacial comum: a informação extra surge então como *pop-up* sobre o ovo, sobrepondo-se a este, tomando as suas coordenadas (Figura 28).



FIGURA 28 – CÍRCULO *POP-UP*, RELATIVO AO OVO QUE FOI CLICADO.

Esta regra não pôde ser cumprida para todas as posições possíveis dentro da visualização: os ovos que ocupavam posições mais periféricas originavam *pop-ups* que mostravam informação fora do ecrã. Para esses casos foi necessário implementar ajustes na posição destes *pop-ups*, de forma a estes não saírem do ecrã. Procurou-se então esbater esta quebra da relação espacial existente recorrendo ao mecanismo de *fade-in*, que permitissem ao utilizador conseguir perceber de onde é que o *pop-up* era originário. Ao apresentar este *feedback* visual seria possível perceber a alteração de estado que foi activada pelo clique.

Para além da manipulação directa dos elementos, existe outro conjunto de ferramentas de interacção que foram implementadas. Estas ferramentas envolvem operações de manipulação da amostra total dos dados. Escolheu-se um espaço comum para agregar todas as ferramentas deste tipo: uma região rectangular na parte inferior da visualização (como se pode ver na Figura 32 em baixo, na página 53). Deu-se a esta região uma cor cinzenta neutra, para não roubar protagonismo visual ao próprio ecrã de

visualização. À volta do ecrã de visualização foi inserida uma moldura com a mesma cor neutra, de modo a se fundir perceptivamente com a região de interacção.

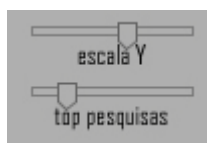


FIGURA 29 – PRIMEIRAS FERRAMENTAS DE INTERACÇÃO A SEREM CONCEPUALIZADAS.

As primeiras funções de manipulação a serem estudadas foram a escala do eixo vertical e o número máximo de elementos a visualizar (o top elementos) (Figura 29). Foi tomada a decisão de utilizar o Slide (*slider*) como ferramenta de interacção para estas funções. Seguindo a mesma ideia de toda a barra inferior de interacção, este *Slide* não tem nenhum elemento visual destacável que roube destaque ao ecrã de visualização.

Posteriormente, o *Slide* relativo à escala do eixo vertical foi retirado, porque foi entretanto alterado o modo de amostragem dos elementos. Antes, o limite máximo vertical era um valor fixo, o que fazia com que a amostra de valores nem sempre se ajustasse da melhor maneira ao eixo vertical. Era então necessário ao utilizador ajustar essa altura máxima através da ferramenta de ajuste da escala, de forma a distribuir melhor a amostra pelo ecrã.

Quanto ao top pesquisas, este foi mantendo praticamente a mesma estrutura ao longo do desenvolvimento. Foi criada outra ferramenta (Figura 30, em baixo) cuja finalidade é semelhante a esta, mas que refina especifica mais a selecção do utilizador. Através de um Slide duplo, é permitida selecção de um mínimo e máximo absoluto de pesquisas que definem o intervalo de selecção de onde se constrói a amostra dos dados. Desta forma, só vão estar visíveis no ecrã palavras-pesquisa cujo número de ocorrências esteja dentro do intervalo definido (maiores e menores que o mínimo e máximo definido, respectivamente). Foi necessário incorporar no espaço entre os Slides de mínimo e máximo um feedback visual que simbolizasse se o intervalo definido retornava uma amostra vazia ou não. Introduziu-se então a cor (vermelho para um intervalo com elementos e verde para um intervalo vazio) para transmitir rapidamente essa informação ao utilizador.

Como estas duas ferramentas originavam a mesma reorganização da amostra, a sua existência conjunta poderia causar alguma confusão, tanto para o utilizador como para o próprio programa. A solução adoptada consistiu na colocação de dois Botões de Radio (*radio buttons*) que permitiam a activação de apenas uma destas ferramentas (Figura 30).

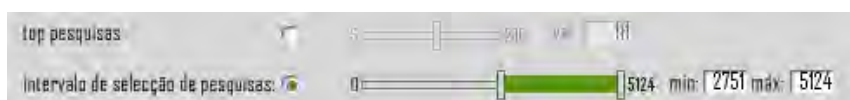


FIGURA 30 – SOLUÇÃO ESCOLHIDA PARA AS FERRAMENTAS PARA SELECIONAR O TOP, OU O INTERVALO DE SELECÇÃO DAS PESQUISAS.

Foram também inseridas caixas de texto, colocadas ao lado dos Slides, com o intuito de fornecerem feedback visual sobre os valores seleccionados pela ferramenta, para guiar o utilizador nesta selecção.

Aproveitando também o acesso a datas associadas às palavras-pesquisa, considerou-se que poderia ser útil para o utilizador definir qual a escala temporal a que pretendia analisar os dados (Figura 31, em cima). Considerando a aplicação tanto para curto como para longo prazo desta análise, implicaria objectivos diferentes a visualização das pesquisas separadas por hora, ou separadas por ano. Foi então desenhado um conjunto de botões que modificassem essa escala, podendo esta ser medida por minutos, horas, dias, meses e anos. Estes botões são semelhantes aos botões de rádio usados nas ferramentas anteriores, na medida em que apenas um pode ser seleccionado ao mesmo tempo, mas com algumas diferenças funcionais dos anteriores (que são explicadas posteriormente na secção do design técnico). Tendo isto em

conta, seria então necessário que o seu *layout* algumas diferenças. Foi usada a ideia visual de criar uma pequena escala conceptual, representada por rectângulos com uma dimensão crescente, representativa da crescente escala (de minutos para anos).



FIGURA 31 – AS DUAS FERRAMENTAS DE INTERACÇÃO: A ESCALA TEMPORAL E A CAIXA DE TEXTO.

Finalmente, foi criada uma caixa de texto que permite ao utilizador localizar, dentro da amostra estabelecida na visualização, a palavra-pesquisa escrita (Figura 31, em baixo). Se esta for encontrada, o sistema marcará o ovo correspondente à palavra preenchendo o ovo com uma cor de destaque. Desta forma, poderá tornar-se mais rápido para o utilizador clicar nesse ovo para ver mais detalhes.

A Figura seguinte (Figura 32) representa o aspecto visual final que foi posteriormente implementado no protótipo.



FIGURA 32 – APARÊNCIA FINAL DA CONCEPTUALIZAÇÃO, ADOPTADA PARA A IMPLEMENTAÇÃO DO PROTÓTIPO.

3.6 DESIGN TÉCNICO

Esta secção refere a implementação técnica do protótipo, procurando explicar o processo de construção das diferentes componentes internas do protótipo, que fazem as suas diferentes partes funcionarem e comunicarem entre elas.

Distingue-se dentro desta secção 3 divisões principais, que reflectem diferentes modelos de programação, tendo objectivos e resultados diferentes: Recolha e tratamento de dados, Visualização e Interacção.

3.6.1 RECOLHA DOS DADOS BROKER

O primeiro passo para todo o processo de construção do protótipo consiste na aquisição e filtragem dos dados através do sistema *Broker* (previamente explicado) que devem ser armazenados e só posteriormente utilizados para a sua visualização.

Os dados são recolhidos para um ficheiro de texto. Após a gravação deste ficheiro de texto, foi construído um programa em *Processing* cujo algoritmo filtra o pacote de dados deste ficheiro, organizando-os numa tabela mais simplificada. Este processo de filtragem percorre linha a linha o ficheiro de texto, e procura *Strings* específicas, que vão delimitar os dados que se pretendem adquirir. Por exemplo temos a linha `<keywords>correio da manha</keywords>`, que surge dentro de um pacote de pesquisa que é delimitado também por outras *Strings* específicas. O código vai comparar a linha com todas as expressões regulares⁵⁴ definidas pelo programador. Uma destas expressões estipula que uma linha com as *tags* `<keywords>` e `</keywords>` deve ter, entre estas duas *tags*, a informação que se deve filtrar.

Esta informação é então guardada como uma *String*, que mais tarde é guardada para um *Array* relativo a esse bloco de pesquisas. O mesmo acontece para as outras *tags* relevantes relativas à data, região e cidade. Após este processo repetir-se ao longo de todo o ficheiro de texto inicial, o programa é concluído com a escrita de um novo ficheiro de texto, desta vez, organizado por linhas representativas de uma pesquisa; cada linha deste ficheiro armazena tanto a frase de pesquisa, como a data da pesquisa, a cidade e a região. Estes valores são separados pelo carácter especial *TAB*, escolhido para facilitar o processo posterior de leitura e separação destes dados⁵⁵.

A próxima fase do processo de tratamento de dados envolve tanto a criação de um top com as palavras-pesquisa mais efectuadas, como a associação destas pesquisas ao número de ocorrências por região do país. Estes dois objectivos de processamento envolvem mecanismos diferentes, e é relevante dividi-los em duas componentes diferentes: Pesquisas e Georreferenciação.

a) PESQUISAS

Esta parte do código diz respeito à criação de uma lista de pesquisas, ordenada por ordem decrescente por número de pesquisas. Começa com o acesso ao ficheiro de texto criado no passo anterior, contendo todas as pesquisas (frases completas de pesquisas) efectuadas. São criados então vários *Arrays* vazios relativos aos diferentes dados a armazenar: palavra-pesquisa, número de ocorrências, data, e as regiões do país (com os nomes: *palavras*, *n_pesquisas* e *datas*, *Aveiro até viseu*). O objectivo é que todos estes *Arrays*, sejam preenchidos simultaneamente, de maneira a que os seus elementos estejam relacionados através de um índice comum. O mesmo poderia ser feito através de um *Array* bidimensional, onde cada coluna corresponderia a uma palavra-pesquisa, e cada linha ao número de ocorrências, data, país, entre outros. Esta solução não foi no entanto adoptada porque foram encontrados alguns problemas na manipulação de *Arrays* desta natureza (na inserção e remoção de colunas), preferindo-se assim seguir por uma opção menos elegante mas mais fiável.

De seguida, é executada uma função iterativa (*contar_palavras_chave()*) que, para cada linha do ficheiro de texto, executa um conjunto de instruções; primeiro, identifica a frase-pesquisa da linha e separa cada palavra desta frase, adicionando-a a um *Array* de *Strings* temporário; os elementos deste *Array* temporário vão ser comparados a cada um dos elementos do *Array* *palavras*, criado anteriormente. Se não for encontrada a palavra correspondente no *Array* *palavras*, esta será inserida no *Array* *palavras* como novo

⁵⁴ Saber mais em: http://en.wikipedia.org/wiki/Regular_expressions

⁵⁵ Poderia ter sido escolhido outro carácter como a vírgula (","), ponto final ("."), etc., mas considerou-se que tal poderia causar problemas, tendo em conta que caracteres deste género poderiam já existir na frase de pesquisa.

elemento; ao mesmo tempo também são inseridos os outros dados correspondentes a esta palavra (data, número de pesquisas, cujo valor por defeito é 1) nos *Arrays* seus respectivos (datas, *n_pesquisas*, entre outros), ligados pelo mesmo índice, o último a ser adicionado aos *Arrays*. Desta forma, sempre que se encontrar uma palavra-pesquisa nova, esta será memorizada, juntamente com os dados a ela relacionados.

Por outro lado, se encontrar uma palavra que já esteja inserida no *Array*, apenas altera os valores relativos ao número de ocorrências (somando mais um), à data (substituindo a última data pela nova, porque de cima para baixo, é a ordem cronológica no ficheiro de texto), e às regiões.

Após finalizar este conjunto de tarefas, é executada uma função que vai ordenar por ordem decrescente, os *Arrays* de dados, consoante o número de ocorrências. Estando todos ligados pelo mesmo índice, a troca dá-se simultaneamente para manter essa relação.

Mais uma vez, esta listagem é exportada para um ficheiro de dados, que guarda na mesma linha os dados dos *Arrays* com o mesmo índice, separados por *TABs*.

b) GEORREFERENCIAÇÃO

O processo de Georreferenciação das pesquisas é alcançado através de um método de comparação directa de *Strings* (neste caso cidades), para determinar qual o distrito a que se pertence. Destaca-se que este pode não ser o método mais fiável a adoptar, tendo em conta que a língua portuguesa é composta de muitos caracteres especiais, podendo originar falhas na comparação.

No entanto, este método foi adoptado pelo facto de que o processo de Georreferenciação do programa pré-existente do *Broker* Sapo não ter sido determinado no decorrer da investigação. Seria de esperar que a *tag* recolhida do *Broker* com o nome <region> representasse, através de um código numérico, a região do país de onde se deu a pesquisa correspondente. Pela troca de e-mails com alguns agentes responsáveis do Sapo, verificou-se que a ideia seria essa, mas que na prática ainda não estava a ser usado esse *tag* para tal operação, e que brevemente se modificaria essa situação. Entretanto não foram dadas mais notificações sobre esta alteração, e foi então necessário criar uma solução alternativa.

Esta solução tem o objectivo de utilizar o campo <city>, que também vem incluído como no pacote de dados do *Broker*, que indica a localidade de onde foi efectuada a pesquisa. Este campo será então usado para verificar qual a respectiva região.

A maior dificuldade esteve em encontrar um ficheiro que já incluísse a listagem de todas as localidades do país e a região a que estas pertencessem. Após algum tempo nesta demanda, teve de se dar primazia aos objectivos principais da investigação; sendo o principal objectivo a criação de um protótipo de visualização onde a ênfase geral da investigação seria a exploração das metáforas visuais, o rigor estatístico não seria nesta fase tão importante.

Criou-se então, também utilizando a ferramenta de *Processing*, um método capaz de criar uma tabela que enumere todas cidades do país e as suas regiões correspondentes. Este programa utiliza dados do sítio web *Ciberforma*⁵⁶, armazenando-os numa tabela (guardada num ficheiro de texto, mantendo a mesma estratégia de linha por linha e separação dos valores por *TAB*). Para o acesso e aquisição automática dos dados também foram utilizadas expressões regulares, de forma a poder-se detectar a listagem dos respectivos concelhos, no código fonte das páginas web do portal, relativas a cada distrito. Por exemplo na página web do distrito de Aveiro⁵⁷, através da leitura do seu código fonte, podemos verificar que a listagem dos concelhos encontra-se com a seguinte nomenclatura:

⁵⁶ Disponível em: <http://codigopostal.ciberforma.pt/>

⁵⁷ Disponível em: <http://codigopostal.ciberforma.pt/Distrito.asp?dist=01>

(...)

```
<area shape="poly" alt="Concelho da Mealhada" coords="119,250, 131,250, 142,258, 183,256, 184,271, 176,275, 164,273, 162,283, 148,285, 144,299, 127,299, 114,295, 108,284, 126,283, 128,267" href="concelho.asp?dis=01&con=11">
```

```
<area shape="poly" alt="Concelho de Anadia" coords="92,229, 123,221, 133,217, 172,218, 178,215, 196,216, 192,220, 198,236, 188,243, 185,254, 173,255, 167,256, 142,257, 132,249, 118,249, 120,241, 114,240, 104,251, 89,250, 89,240" href="concelho.asp?dis=01&con=03">
```

(...)

É possível então filtrar todo o código que envolve o nome do concelho, que é a informação que se pretende retirar. Através de expressões regulares retira-se todo o texto desnecessário que engloba o que pretendemos, adquirindo-se apenas o nome do concelho; parte-se então do princípio que as outras páginas *web* deste portal têm a mesma estrutura no código fonte, permitindo ao programa utilizar o mesmo conjunto de instruções. Também foi necessário aferir se o URL correspondente a cada página de distrito teria também um padrão na sua nomenclatura, que permitisse ao programa aceder-lhes automaticamente. Verificou-se que apenas se alterava este URL nos últimos dois dígitos, sendo a restante parte do URL idêntica para todos os distritos. Como estes dois dígitos são numéricos, foi possível concatenar a parte comum do URL com a variável contadora do ciclo.

Como resultado final deste programa, temos uma lista que associa directamente cidades a distrito. Uma das grandes limitações é o facto das pesquisas que não forem efectuadas em cidades-concelho passarem não têm referência geográfica, não tendo posteriormente lugar no mapa geográfico.

O ficheiro de texto criado vai então ser consultado pelo programa que vai criar o top das pesquisas, referido anteriormente.

Os *Arrays* criados para as regiões são chamados, sendo cada um deles relativo a uma região diferente do país. Tal como anteriormente, quando o programa detectar uma nova palavra-pesquisa, vai procurar a cidade a que corresponde essa palavra e determinar (utilizando o ficheiro de texto dos distritos) qual o distrito correspondente, através de uma função *calcula_distrito()*. Se encontrar correspondência, insere um novo elemento no *Array* do distrito (Aveiro, ou Beja, ou Braga, por exemplo) que foi encontrado, ligado pelo índice ao *Array palavras*, cuja actualização é dada na mesma altura. Se, pelo contrário, a palavra-pesquisa já se encontrar inserida neste *Array*, o programa vai somar mais 1 ao elemento do *Array* do distrito, correspondente ao *Array palavras* (Tabela 1).

TABELA 1 – REPRESENTAÇÃO ESQUEMÁTICA DO CONJUNTO DOS ARRAYS RELATIVOS AOS DADOS DE PESQUISA

Índice: Array:	0	1	2	...
<i>palavras</i>	"google"	"portugal"	"futebol"	...
<i>datas</i>	20091025145012	20091001094024	20091105131200	...
<i>n_pesquisas</i>	23	52	10	...
<i>Aveiro</i>	3	10	0	...
<i>Beja</i>	0	1	1	...
...

Na tabela 1 estão representados em cada linha os *Arrays* relativos aos dados registados. Cada coluna está associada a um índice concreto, que por sua vez está relacionado com uma palavra-pesquisa única, que é especificada pelo *Array palavras*. Analogamente, sempre que uma palavra-pesquisa nova é identificada, que não está presente nesta tabela, é criada uma nova coluna, o que implica que para cada linha (*Array*) seja criada uma nova entrada com o mesmo índice dessa coluna.

Por fim, quando se executa a função de ordenação dos *Arrays* por número de ocorrências (*ordenar_frequencia*), está-se basicamente a ordenar a ordem das colunas pela ordem das colunas da linha *n_pesquisas*; isto implica também a troca das colunas pelas ao longo das linhas (*Arrays*) restantes, de modo a permitir que estas mantenham a sua relação.

Após todos estes processos serem concluídos, é finalmente executada o comando que irá escrever o conteúdo dos *Arrays* para o ficheiro de texto de *output*, onde cada índice (ou coluna) será escrito numa linha diferente do ficheiro.

O ficheiro de texto de saída terá os dados aproximadamente no seguinte formato (Tabela 2):

TABELA 2 – FORMATO DO FICHEIRO COM O TOP DE PESQUISAS

portugal	52	20091001094024	10	1	20	0	0	1	...
google	23	20091025145012	3	0	5	0	0	0	...
futebol	10	20091105131200	0	1	0	1	0	4	...

Onde a partir da terceira coluna surgem os dados relativos às regiões por ordem alfabética (Aveiro, Beja, Braga, etc.). A separação das colunas é representada pelo carácter *TAB*.

Este ficheiro de texto será a principal fonte de dados a ser utilizada no programa de visualização.

3.6.2 VISUALIZAÇÃO

Após a recolha e formatação dos dados, é inicializado o programa que vai dar forma a estes dados, ao mesmo tempo que disponibiliza ferramentas de interacção para manipular a visualização. Também faz sentido dividir o design técnico deste programa em duas partes distintas – Pesquisas e Geovisualização, sendo duas componentes diferentes, que são mais tarde conectadas através da acção do utilizador (clique de rato).

a) PESQUISAS

A representação de pesquisas envolve um certo número de rotinas onde é importante explicar as mais importantes.

Inicialmente, é acedido o ficheiro que inclui o top das pesquisas criado anteriormente (com a estrutura do exemplo da Tabela 2). Como o valor por defeito de pesquisas a visualizar é igual a 100 (definido pelo programador), o programa quando se inicia vai criar 100 instâncias da classe bola, que representativas das 100 pesquisas mais efectuadas; isto implica que o programa realize operações com dados das 100 primeiras linhas do ficheiro de texto⁵⁸. Linha a linha, vai usar os dados recolhidos para criar um novo elemento do *Array* de classes bola (explicado mais à frente). Cada instância que é criada vai embeber dentro de si estes dados para se modelar de acordo com estes. Com base no número de pesquisas a que corresponde, a bola (ovo) será situada numa altura proporcional, assim como terá um tamanho relacionado a este valor, mas de um modo mais indirecto. Este tamanho está mais relacionado com a percentagem de pesquisas efectuadas relativa à amostra de dados seleccionada; isto quer dizer que mesmo antes de se inicializarem as bolas, é necessário ter uma contagem preliminar do total de pesquisas da amostra de dados (que na inicialização do programa são por defeito as 100 primeiras palavras mais pesquisadas).

O procedimento que efectuará esta contagem estará também presente sempre que se efectuarem alterações na amostra da visualização, pois tais operações resultarão sempre na reformulação das bolas, com novos parâmetros de entrada.

É necessário perceber o que é que as instâncias criadas nesta altura significam, sendo fulcral perceber a classe de onde estas derivam.

⁵⁸ Na verdade, tal repete-se até à linha 101, pois a contagem começa na 2ª linha do ficheiro de texto, pois a primeira linha encontra-se vazia. Este formato do ficheiro de texto foi escolhido para contrariar erros que aconteciam em algumas comparações de Strings; as que se encontravam na primeira linha do ficheiro eram sempre consideradas diferentes com o seu equivalente noutras linhas.

CLASSE “BOLA”

A classe *Bola* é o “corpo” dos ovos da visualização, que representam as pesquisas efectuadas, e aglomeram todas as características necessárias para o seu funcionamento e comunicação com o programa principal.

Recebe, aquando a criação de uma instância sua, os seguintes parâmetros de entrada: um número de identidade (ID), único para cada bola; a coordenada x e y da sua posição inicial (que no início é aleatória, dentro dos limites do ecrã); o número de pesquisas efectuadas; a palavra-pesquisa a que corresponde; a data da última pesquisa efectuada; o *Array* total das instâncias da classe (utilizado para a comunicação entre as várias instâncias); por fim, um *Array* agregador dos valores de cada região do país, posteriormente usado para a Georreferenciação.

Na sua inicialização, são executadas automaticamente operações de *setup*, que são executadas apenas uma vez. Estas operações servem para converter todas as variáveis de entrada em variáveis internas da classe; também no *setup* é convertido o número de pesquisas na percentagem relativa de pesquisas, que será convertida no tamanho do ovo – esta conversão volta a ser realizada na própria função de desenho da bola.

Existem 3 funções especialmente importantes nesta classe, que são chamadas externamente, pelo corpo principal do programa; a função *display()*, a função *move()* e a função *mouseClicked()* representam as características base dos ovos – a sua aparência, o seu movimento, e a sua interactividade, respectivamente, e é através destas que são executadas todas as outras funções internas da classe bola.

A função *display()* da classe bola é utilizada para estar em constante ciclo de execução, e é a responsável por desenhar no ecrã a representação visual desta instância; é chamada na função *draw()* do corpo principal do programa, na medida o desenho dos ovos está em permanente execução.

A função *move()* também está em constante execução, sendo responsável pela variação da posição da bola ao longo da execução do programa, posição esta ditada através de equações de movimento.

Por fim a função *mouseClicked()* apenas é accionada quando ocorrem eventos de clique do rato, e se a posição do rato se encontrar dentro dos limites do ovo, é activado um conjunto de instruções que tanto vão influenciar eventos na função *display()* como as restantes instâncias de *Bola*.

Estas 3 funções estão interligadas, na medida em que cada uma pode alterar variáveis internas da classe que vão alterar o algoritmo nas restantes. Por exemplo, se um ovo verificar que o rato clicou de facto dentro dos seus limites (através da função *mousePressed()*), este parará de se mover, pois alterará a função *move()*; ao mesmo tempo, activará na função de *display()* uma máscara que mudará a sua cor, indicando que esta está em destaque. A função *display()* também pode neste caso comunicar exteriormente, para o corpo principal do programa, transmitindo o ID da bola, para que este saiba que tem de desenhar mais detalhes, sobre essa bola.

De seguida explicam-se mais detalhadamente estas 3 funções.

BOLA.DISPLAY()

A função *display()* da classe bola tem uma série de procedimentos que executa por ordem sequencial, sendo alguns deles especialmente importantes para se perceber a dinâmica desta classe.

Uma das primeiras operações realizada nesta função, é a redefinição dos valores para as variáveis x0 e y0, incluídas nas equações de movimento da bola.

A variável x0 representa a posição no eixo horizontal onde a bola deve encontrar a sua posição de equilíbrio, ou seja, onde esta deve ficar posicionada com velocidade horizontal nula. Esta variável está dependente da atribuição de uma escala temporal ao eixo horizontal; se nenhum botão da escala temporal tiver sido activado, e nenhuma escala ter sido atribuída, a posição de equilíbrio horizontal tomará um valor

O círculo de *pop-up* é desenhado no corpo principal do programa, através da função *desenha_destaque()*. É executada externamente à classe para que se desenhe o círculo de *pop-up* sobreposto sobre todos os outros ovos. Esta comunicação com o corpo principal do programa dá-se sobretudo através da mudança de valores de variáveis globais, que são uma ferramenta útil de comunicação entre módulos, pois são reconhecidas por qualquer componente do sistema.

BOLA.MOVE()

Sendo esta uma função particularmente simples, interessa perceber como é que se dá o movimento dos ovos, e como é que este movimento pode parecer fluído e não artificial. Também esta função é executada da mesma maneira que a função *display()*, na medida em que está em constante execução no corpo principal do programa.

Existem 3 problemas principais que esta função aborda; o primeiro está relacionado a flutuação dos ovos, evitando uma aparência estática quando numa posição de equilíbrio; o segundo diz respeito à movimentação que se dá para essa situação de equilíbrio; por sua vez, o terceiro problema trata das limitações a impor aos ovos para que estes se mantenham dentro dos limites do ecrã.

A função *move()* começa por verificar se o ovo em questão foi clicado (*em_cima*): se tal aconteceu, a função não executa nenhuma instrução, mantendo-se parada; se esta condição se verificar falsa, a função executa o seu conjunto usual de operações.

Na abordagem ao primeiro problema, referido anteriormente, é adicionado um factor vibratório à posição horizontal do ovo, para que este oscile ligeiramente neste eixo; apesar desta oscilação, o ovo deve manter-se próximo da sua posição de equilíbrio horizontal (x_0); isto é operacionalizado através da verificação constante da sua posição horizontal em relação a um intervalo definido por x_0 menos 2 pixels e x_0 mais 2 pixels ($[x_0-2; x_0+2]$). Ou seja pretende-se que o ovo não oscile para além de 2 pixels da sua posição de equilíbrio. Enquanto se mantiver dentro deste intervalo, é atribuído à variável de vibração um valor aleatório entre -0.05 e 0.05. Esta variável de vibração será posteriormente adicionada á equação de movimento do ovo.

Quando se verificar que a bola atingiu um dos extremos deste intervalo ($[x_0-2; x_0+2]$), a variável de vibração tomará outro valor, de forma a afastar o ovo do limite atingido; por exemplo, se a bola tiver atingido o limite inferior (x_0-2), a variável de vibração tomará um valor aleatório entre 0 e 0.05, e se atingir o limite superior (x_0+2), tomará um valor aleatório entre -0.05 e 0.

O segundo problema diz respeito à própria movimentação dos ovos ao longo do ecrã de visualização. Como já temos calculadas as posições de equilíbrio (y_0 e x_0), resta associa-las à posição actual do ovo, através das seguintes equações de movimento (Figura 34):

$$\begin{aligned} x &= x + vib + \frac{x_0 - x}{constante} \\ y &= y + \frac{y_0 - y}{constante * 0.7} \end{aligned}$$

FIGURA 34 – EQUAÇÕES DE MOVIMENTO (HORIZONTAL E VERTICAL) DOS OVOS

Na Figura 34 as variáveis x e y representam a posição actual do ovo, vib a variável de vibração, e *constante* representa uma variável de valor fixo, usada para abrandar a velocidade de deslocação das bolas para o seu ponto de equilíbrio. Neste caso toma o valor de 10, sendo esta atribuição apenas baseada na escolha pessoal do programador, na sua percepção de qual a velocidade mais indicada dos ovos. Repare-se que este valor para o eixo vertical é multiplicado por 0.7, o que significa que a deslocação neste eixo será ligeiramente mais lenta que no eixo horizontal.

O que estas equações mostram visualmente é que quanto mais próximo estiver o ovo da sua posição de equilíbrio (x_0 e y_0), menor será a sua velocidade de aproximação; esta velocidade tende a nula com o tempo: quando o ovo chegar à posição de equilíbrio, x_0 será igual a x , e y_0 será igual a y .

Por fim o terceiro problema mencionado, o da limitação dos ovos dentro do ecrã de visualização, esse problema é corrigido na própria definição dos valores de x_0 e y_0 , na função *display()*, que já são mapeados de maneira a tomarem apenas valores dentro dos limites da visualização. Isto implica que nunca irão surgir pontos de equilíbrio que requeiram aos ovos a sua movimentação para fora do ecrã.

BOLA.MOUSECLICKED()

Esta função é chamada pelo corpo principal do programa, dentro da sua função *mouseReleased()*. A função *mouseReleased()* é executada automaticamente sempre que se verifica um evento de clique de rato dentro da aplicação. Dentro desta encontram-se em execução todas as funções de interacção envolvendo o clique de rato, para qualquer elemento interactivo na aplicação. Uma destas funções é a função *mouseClicked()* da classe *Bola*, que através de um ciclo é executada para cada instância dessa classe.

Dentro da função *mouseReleased()* de cada ovo é comparada a posição do rato com a posição desse ovo, de maneira a saber se quando é clicado o botão do rato, este se encontra em cima de algum ovo em específico.

Se esta condição for verdadeira, altera-se o valor da variável booleana *em_cima*, que vai transmitir ao longo da instância que esta foi clicada, de modo a que realize operações em concordância com este facto. No entanto, o que se vai fazer não é atribuir a esta variável o valor verdadeiro, mas sim atribuir o oposto do valor que esta já tem: *em_cima = !em_cima*;

Esta atribuição, apesar de simples, é importante na medida em que permite que um segundo clique sobre o mesmo ovo anteriormente clicado desligue o círculo de *pop-up* que fora activado, e volte à condição inicial.

Outra instrução importante ocorrida ainda dentro da função *mouseClicked()* é a atribuição do valor falso para as variáveis *em_cima* das restantes instâncias da classe (os restantes ovos). Desta forma, previne-se que seja possível a coexistência simultânea de vários círculos de *pop-up* na visualização, tornando-a confusa.

Se por outro lado, o cursor não se encontrar sobre a instância quando o rato for clicado, a função executa um conjunto pequeno de procedimentos de precaução. Estes procedimentos podem ser a atribuição do valor falso à variável *em_cima* e o *reset* para zero da variável *timer* (utilizada para efeitos de *fade-in* e *fade-out*).

b) GEORREFERENCIAÇÃO

No que diz respeito à construção do mapa regional relativo à percentagem de pesquisas para cada palavra, este processo começa fora do *Processing*. Foi recolhida uma imagem de Portugal continental, já separado por distritos⁵⁹. Esta imagem foi então transformada usando a ferramenta de edição de imagens *Photoshop*. Para além de se efectuarem os ajustes da imagem relativos ao seu aspecto visual, também se separaram os distritos desenhados em camadas diferentes (*layers*) do ficheiro da imagem. Recolheu-se então a posição, em x e y , de cada uma destas imagens, relativamente ao mapa total. De seguida, cada uma dessas camadas foi guardada em ficheiros diferentes. Cada uma destes ficheiros de imagem, representa uma peça do mapa do país, que serão reagrupados posteriormente pela aplicação.

⁵⁹ Retirada de: http://upload.wikimedia.org/wikipedia/commons/7/79/Mapa_de_Portugal_-_Distritos_plain.png

Já no programa escrito em *Processing*, o processo de reagrupamento começa com o *upload* destes ficheiros de imagem para a memória do programa, assim como de um ficheiro de texto, contendo a listagem das posições para cada uma dessas imagens; este ficheiro é escrito pelo programador, e segue a mesma tipologia de linha a linha, com os dados separados por *TABs* (distrito, x, y).

Por fim é criado um *Array* da classe *Regiao*, que agrega 18 instâncias desta classe, representativas de cada região de Portugal continental.

CLASSE *REGIAO*

Esta classe é inicializada com as seguintes variáveis de entrada: um número inteiro identificativo da região (ID), as coordenadas x e y recolhidas do ficheiro de texto (x e y), a imagem associada à região, e uma *String* indicativa do nome da região representada pela instância.

Sendo esta classe representativa de cada região do país (apenas Portugal continental), o *Array* criado das suas instâncias tem um número fixo (18) que vai permanecer fixo ao longo do programa.

Quando uma instância é criada, corre uma série de instruções de instalação, de cópia das variáveis de entrada para variáveis internas à classe, que só ocorrem uma vez.

REGIAO.DISPLAY()

A função mais relevante dentro desta classe é a função *display()*. Esta função agrega todos os procedimentos que vão levar à representação visual da região, tendo em conta o número de pesquisas efectuadas a determinada palavra. Começa com uma sub-função de nome *calcula_percentagem()*, que como o nome indica calcula o valor percentual de pesquisas, em relação ao total das pesquisas. Esta sub-função utiliza como parâmetro de entrada a variável *n_pesquisas* (cujo valor por defeito é zero). O valor desta variável é definido na função *desenha_destaque()*, executada no corpo principal do programa, que é responsável pelo desenho de todo o círculo de *pop-up*, incluindo o desenho das regiões.

De seguida, com base no valor retornado pela função *calcula_percentagem()*, vai-se inquirir em que intervalo de valores se insere este valor.

Estes intervalos representam percentagens que vão inserir os valores dentro de intervalos, que vão representar cores diferentes no mapa; são estabelecidos pelo programador, e armazenados num *Array* global, permanecendo imutáveis ao longo do programa. Neste caso, foi utilizada um conjunto de valores que seguem uma fórmula específica (Figura 35):

$$val = \frac{75}{n^2}$$

FIGURA 35 – SÉRIE GERADORA DA ESCALA PERCENTUAL DO NÚMERO DE PESQUISAS GEORREFERENCIADAS.

Os valores de da variável *n* representada na fórmula da Figura 35 vão de 1 até 5, definindo assim 5 intervalos distintos. Os valores resultantes são 75.0%, 18.75%, 8.33%, 4.69% e 3.0%.

Em vez de uma série linear de dados (de por exemplo: 100%, 75%, 50%, 25%, 0%), optou-se por usar esta gama de valores. Foi escolhida com base na interpretação subjectiva do programador sobre qual a gama de valores que seria a mais adequada à distribuição de valores, conseguindo representar maiorias relativas e absolutas dentro desta distribuição.

Com base na comparação entre o valor percentual de pesquisas e os intervalos de valores definidos ([100;75], [75;18.75], [18.75;8.33], [8.33;4.69], [4.69;3]), é atribuído um valor diferente da opacidade (*alfa*) da cor, a ser utilizada na coloração da imagem da região (sendo estes valores: 250, 200, 150, 100, 50). A tonalidade da cor permanece a mesma (verde), sendo a opacidade o factor de distinção de regiões com percentagens diferentes.

De seguida, a função *display()* vai proceder ao desenho da região, utilizando um efeito de *fade-in*; a imagem da região é desenhada na sua posição pré-definida, com a coloração definida anteriormente, assim como escrever sobre esta o valor da percentagem correspondente.

É a função *desenha_destaque()* que comanda esta acção, na medida em que chama de um modo sequencial, a função *display()* de cada uma das instâncias da região.

DESENHA_DESTAQUE()

Esta função é responsável pelo desenho do círculo de *pop-up*, que aglomera o mapa regional das pesquisas, a palavra-pesquisa, o número de pesquisas, a percentagem relativa de pesquisas, a data da última pesquisa efectuada, e também a percentagem de pesquisas georreferenciadas; este último valor é calculado em relação ao número total de pesquisas da palavra-pesquisa, calculando a percentagem das que estão relacionadas a uma região; é um indicador importante, pois traduz o nível de representatividade que o mapa geográfico desenhado tem sobre o total de ocorrências da palavra em destaque.

Como parâmetro de entrada, a função *desenha_destaque()* tem apenas uma variável, que representa a instância da classe *Bola* da qual se pretende ver os destaques; este valor é enviado pela própria instância, quando é clicada, através de uma variável global do sistema (*i_bola*).

A função vai usar este valor para adquirir da instância de *Bola* correspondente as suas coordenadas x e y, das quais vai ser desenhada em seguida uma bola branca delimitadora do conteúdo que vai ser mostrado de seguida.

É então imposta uma condição que previne que esta bola passe para fora dos limites do ecrã; tal pode acontecer quando esta se desenha a partir da posição de um ovo que esteja situado nos limites do ecrã; a bola, adoptando, a mesma posição central que o ovo, vai mostrar-se parcialmente fora do ecrã, pois tem uma dimensão maior. Esta condição limitadora funciona definindo que toda a área da bola deve permanecer no interior do ecrã de visualização – utilizando a função *constrain()* do *Processing*.

De seguida dar-se-á o conjunto de instruções que vão desenhando as regiões do país, num ciclo que vai percorrer todo o *Array* das instâncias da classe *Regiao*; para tal atribui-se à variável interna à classe *Regiao* do número de pesquisas o valor que está inserido na instância da classe *Bola* em questão; ajusta-se a posição desta região tendo em conta a posição central do círculo de *pop-up*, e desenha-se a imagem, chamando a função *display()* da classe *Regiao*.

São desenhadas as legendas do mapa, cujo tamanho e posição relativa ao círculo envolvente tem um valor fixo. Por fim, mostra-se a informação restante: a última pesquisa efectuada, a palavra-pesquisa, o número de pesquisas, por exemplo.

A (Figura 36) ilustra todo o processo de desenhar o círculo de *pop-up* sobre um ovo de pesquisa, desde o clique do rato até à execução da função no corpo do programa.

Este processo começa com a constante verificação da execução de algum clique dentro da aplicação, responsável pela função *mouseClicked()* do programa principal (1). Se existir um clique (2), a mesma função executará um ciclo que circulará todos os elementos do *Array* das instâncias da classe *Bola* (3). Para cada execução deste ciclo, executar-se-á a função *mouse_clicked()*, interna a cada instância, que verificará se foi a presente instância a ser clicada (4). Se esta verificação for verdadeira, a função da instância *display()* modificará a aparência da bola (5), atribuirá à variável global *id_bola* o valor do ID da instância (6), e indicará à função *move()* da mesma para executar a instrução de parar a bola (7). Por fim será executada a função *desenha_destakes()* do programa principal, cujo parâmetro de entrada é o valor da variável *id_bola* (8).

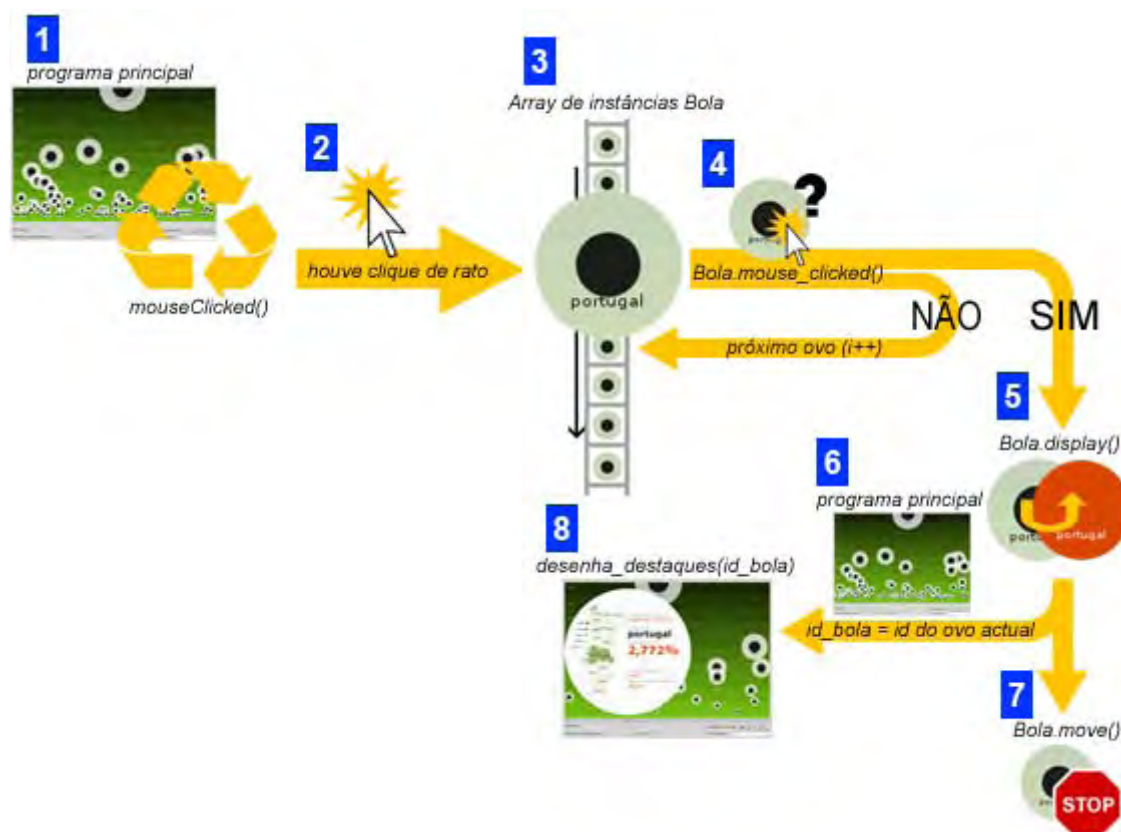


FIGURA 36 – PROCESSO DE EXECUÇÃO DO CÍRCULO DE POP-UP DE UM OVO.

c) FERRAMENTAS DE INTERACÇÃO

Esta secção infere sobre a implementação das ferramentas de interacção do sistema, na sua representação, na maneira como estas alteram as definições da visualização e comunicam essas alterações. Estas ferramentas estão situadas na barra inferior do sistema e são: *Slider* de top de pesquisas, *Slider* duplo de intervalo de pesquisas, botões da escala temporal, os botões de selecção dos segmentos temporais, e a caixa de texto de selecção de pesquisas.

O círculo de *pop-up* (explicado na secção *Desenha_destaque()*, página 63) poderia ser definido como uma ferramenta de interacção, na medida em que surge após a acção directa do utilizador. Porém, este funciona de modo diferente relativamente às restantes ferramentas, estando mais directamente ligado aos ovos.

Quanto às restantes ferramentas de interacção, é desde já relevante esclarecer de que forma é que estas funcionam na generalidade, sendo executadas de maneira semelhante.

Têm normalmente uma classe associada, que contém duas funções internas principais: *display()*, e *mousePressed()* (ou *mouseClicked()*); a primeira é chamada posteriormente na função *draw()* do programa, e a segunda é executada em funções de eventos de rato do programa (*mousePressed()* ou *mouseReleased()*).

SLIDER TOP PESQUISAS (CLASSE *DRAG*)

FIGURA 37 – TOP PESQUISAS.

Esta ferramenta (Figura 37) é responsável pela redefinição da amostra de pesquisas na visualização, pois permite ao utilizador seleccionar quantos elementos com maior número de pesquisas pretende mostrar. De um modo prático, selecciona até que linha do ficheiro de *input*, contendo os dados das pesquisas, se pretende visualizar, podendo esta selecção ir de 5 até 200 pesquisas.

É na função de *Setup* do programa, que é apenas executada na sua inicialização, que vai ser criada uma instância desta classe; quando é criada, esta recebe vários parâmetros de entrada: a posição do botão de *Slider* no programa, a variável que a classe pretende alterar (o número total de ovos, representado pela variável global *total_bolas*), e os valores limite que esta variável pode tomar (5 e 200, para este caso).

Dentro da classe, após correr o *setup* inicial para transferir as variáveis de entrada para variáveis internas, temos as funções *display()*, *mousePressed()*, e *mouseReleased()*.

A função *display()* vai desenhar o botão do *Slider*, que será o objecto que se pode arrastar da ferramenta; esta é a execução por defeito da função *display()*, existindo mais um conjunto de operações que se efectuam quando o *Slider* é clicado: dá-se a atribuição da posição horizontal do rato (*mouseX*) à própria posição *x* do *Slider*; ao mesmo tempo, limita-se esta posição, através da função *constrain()*, aos limites da barra de deslize do *Slider*. De seguida, é convertido o valor de *x* do *Slider*, em função dos seus limites, para um novo valor relativo ao total das bolas (*total_bolas*).

É através das funções *mousePressed()* e *mouseReleased()* que a classe verifica a condição de ter sido clicada: a primeira função afere se a posição do rato se encontra nos limites do *Slider*, atribuindo o valor verdadeiro à variável *em_cima*, enquanto a segunda volta a atribuir a esta variável o valor falso, e comunica ao programa principal reiniciar a criação da amostra de ovos, pois uma nova selecção acabou de ser feita.

SLIDER DUPLO INTERVALO DE PESQUISAS (CLASSE *DOUBLE_DRAG*)

FIGURA 38 – INTERVALO DE PESQUISAS.

Esta ferramenta (Figura 38) tem uma utilidade semelhante à anterior, na medida em que determina o tamanho da amostra sobre a qual se vai criar a visualização; a diferença é que pode definir qualquer intervalo dentro do ficheiro de texto, através da definição do número de pesquisas mínimas e máximas; permite assim visualizar grupos de pesquisas restritos apenas a um intervalo definido de ocorrências.

A sua classe correspondente funciona de modo análogo à classe anterior, com a grande diferença de ter dois *Sliders* a serem desenhados, que alteram parâmetros diferentes da amostra. Por outro lado, incorpora da mesma maneira a função *display()*, *mousePressed()* e *mouseReleased()*, cujo funcionamento é semelhante.

A função *display()* tem agora dois *Sliders* para desenhar por defeito, para além da variável *em_cima*, poder ter 3 valores distintos: 0, 1 e 2; o valor zero representa que nenhum dos *Sliders* foi arrastado, e nenhuma nova operação é requerida; o valor 1 representa que o *Slider* da esquerda foi arrastado, e que se alterou o valor mínimo do intervalo da selecção, enquanto a sua posição segue a do rato. Da mesma forma, quando a variável *em_cima* é igual a 2, quer dizer que o *Slider* da direita foi arrastado, e que este deve seguir a posição horizontal do rato, assim como mapear o valor máximo do intervalo de pesquisas.

É a função *mousePressed()*, que atribui estes 3 valores à variável *em_cima*, baseando-se na posição do rato quando se carrega o botão.

A função *mouseReleased()*, activada quando o botão do rato é libertado, verifica se *em_cima* é igual a 1 ou a 2, e se tal for verdade, transmite para o programa principal que se deve proceder à recriação da amostra com base no novo valor mínimo e máximo, indicados pela posição dos *Sliders*.

BOTÕES DA ESCALA TEMPORAL (CLASSE *TRADIO_BUTTON*)

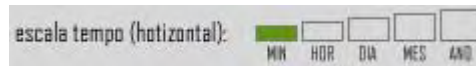


FIGURA 39 – ESCALA TEMPORAL.

Esta ferramenta é constituída por vários botões que definem a escala temporal a aplicar ao eixo horizontal da visualização, activando a distribuição dos ovos ao longo deste eixo consoante o seu valor respectivo (Figura 39).

De um modo operacional, é criado um *Array* de instâncias da classe *TRadio_button* de tamanho definido, no arranque do programa. A criação destas instâncias aceita algumas variáveis de entrada: um número de identidade único, associado ao índice do *Array* das instâncias, as posições x e y para onde se deve desenhar os botões, as dimensões com que se vão desenhar os botões (largura e altura), e o próprio *Array* das instâncias, usado para comparar os botões, entre si.

Após o *setup* inicial de cópia dos parâmetros de entrada para variáveis locais, temos duas funções: *display()* e *mouseClicked()*. A primeira função é relativamente simples: desenha um rectângulo verde, nas coordenadas e dimensões retiradas nas variáveis de entrada da classe, se este botão tiver sido clicado (variável *checked* tem o valor verdadeiro).

É na função *mouseClicked()* que se vão executar as operações mais importantes. Esta começa com a verificação habitual do clique sobre os limites do botão. Se esta condição for verdadeira, atribui-se à variável local *check* o valor contrário ao seu valor existente (*check = !check*). Esta medida é necessária, na medida em que é possível desseleccionar um botão pelo clique de rato nele próprio. Outra forma de o desseleccionar sucederá se o utilizador clicar num botão diferente desta ferramenta; tal é executado através da atribuição do valor falso à variável *check* de todos os outros botões que não foram clicados.

Seguidamente, temos um conjunto de condições, apenas executadas se o valor de *check* for verdadeiro. Estas condições comparam o número de identificação da instância, colectado como variável de entrada da classe. Consoante o valor que esta variável tiver, serão atribuídos valores máximos e mínimos para a escala correspondente. Por exemplo, para o caso do botão da escala das horas, cujo ID é 2, o limite mínimo e máximo será definido como 1 e 24, respectivamente. Convém referir que estes valores são guardados em variáveis globais, sendo mais tarde utilizados, tanto pelo programa principal, como pela classe *Bola*.

Ainda dentro da função *mouseClicked()*, será executada a função de nome *start_eixo_xx()*, que desenha os valores da escala ao longo do eixo horizontal – que representam outra ferramenta de interacção utilizados para seleccionar instantes temporais específicos.

Ao mesmo tempo, as instâncias da classe *Bola* recebem a indicação de que um destes botões foi clicado, e que agora o eixo horizontal já deve ser tido em conta. O valor da posição horizontal dos ovos passa de um valor aleatório a um valor respectivo à data da última pesquisa efectuada, mais especificamente à escala que foi escolhida (ano, mês, hora, minuto).

BOTÕES DE SEGMENTOS TEMPORAIS (CLASSE *TINTERVALO*)

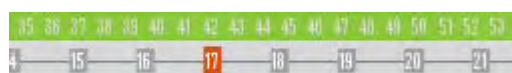


FIGURA 40 – SEGMENTOS TEMPORAIS.

Estas ferramentas de interacção são iniciadas pela acção dos botões anteriormente referidos, e não estão constantemente visíveis na interface do programa. Permitem ao utilizador ver em detalhe as pesquisas relativas só ao instante temporal correspondente ao botão. Por exemplo, se tivermos uma distribuição de pesquisas, alinhados por hora da última pesquisa, e vemos um grupo de ovos alinhados verticalmente sobre as 5h, podemos clicar onde diz “5h”, para vermos só essa amostra de dados, ocultando as restantes que não estão nas 5 horas; o que sucede é que esta sub-amostra é redistribuída pelo eixo horizontal, desta vez adoptando a escala temporal inferior à seleccionada anteriormente. Voltando ao exemplo, a amostra das 5h, quando clicado o botão, vai ser distribuída horizontalmente pelos minutos dessa hora. Esta nova disposição é suportada pelo desenho de uma legenda de valores, relativa aos minutos, sobre a escala das horas; o clique sobre o mesmo botão das 5h retornará a totalidade da amostra à disposição anterior, voltando a mostrar os elementos anteriormente ocultados.

Aquando da sua criação, as instâncias desta classe admitem as seguintes variáveis de entrada: o ID relativo ao seu índice no *Array* das instâncias, a sua posição x, a posição y que é uma constante, e o próprio *Array* das instâncias desta classe, utilizado para comunicar com as outras instâncias.

Convém mencionar que a posição x é determinada pelo mapeamento dos valores do índice do *Array* (que no exemplo das 5h, vai de 1 até 24), distribuídos equitativamente pela largura do ecrã. O valor y vai ser uma constante, pré-definida pelo programador.

Esta classe tem duas funções fulcrais: a função *display()* e a função *mouseClicked()*. Também inclui uma função *mouseOver()*, que apenas transmite instruções para mudar a aparência dos botões caso o rato passe por cima destes, sem no entanto os clicar.

A função *display()* desta classe é bastante extensa: o código mais complexo que a constitui diz respeito à animação que resulta do clique no botão, do “desdobrar” da amostra que está alinhada no instante temporal seleccionado.

Usando o exemplo anterior, com o instante temporal das 5h, podemos sumariar o que se desenrolaria nesta função: após a verificação de ter sido clicada a instância 5, todas os ovos que não tiverem na sua hora de última pesquisa efectuada o valor 5, serão ocultados – isto dá-se através da modificação de uma variável interna à classe *Bola* (*visible*). De seguida dar-se-á uma nova alteração nos valores extremos usados na distribuição dos ovos pelo eixo das horas. Esta alteração passará de 1 e 24 (os extremos relativos às horas) para 1 e 60 (relativos aos minutos, neste exemplo); estes valores extremos são variáveis globais, sendo reconhecidas automaticamente pela classe *Bola* – os ovos saberão assim automaticamente para onde se dirigirem, consultando os minutos da sua data de última pesquisa.

Nesta altura, apenas o botão das 5h poderá fazer com que a amostra volte à sua visualização inicial, mostrando a distribuição de toda a amostra por hora de pesquisa; quando este é clicado, a função *mouseClicked()* da classe atribuirá à variável do clique outro valor (*clicked* igual a 2) que transmitirá à função *display()* a instrução para realizar as operações inversas, tanto de animação como de cálculos. Os ovos voltam assim a usar os valores extremos relativos às horas (neste exemplo, 1 a 24) e a serem dispostos no eixo horizontal com base neste valor; por outro lado os ovos que antes tinham sido ocultados por não serem pesquisados em último às 5h, voltaram a estar visíveis, alinhados com a sua hora correspondente.

FILTRAGEM POR PALAVRA-PESQUISA (CAIXA DE TEXTO)



FIGURA 41 – FILTRAGEM DA PALAVRA-PESQUISA

Esta ferramenta permite a localização de uma determinada palavra de pesquisa, dentro da amostra de visualização, e é constituída por uma caixa de texto, onde o utilizador pode escrever a palavra que quer pesquisar. Se esta palavra for encontrada dentro do conjunto dos elementos no ecrã, programa marcará o ovo correspondente com uma cor de destaque.

A construção desta ferramenta é a única que recorre a uma livreria externa, de uso livre, que fornece ferramentas de interface gráfico já programadas: *Interfascia*⁶⁰. Desta livreria foi apenas utilizado o conjunto de classes relativos à ferramenta de caixa de texto.

Os seus procedimentos são executados no *setup* do programa principal. O local onde se pode definir como é que se pode utilizar o texto que é teclado é na função *actionPerformed()*, que é executada quando se escreve dentro da caixa de texto. Nesta função, o comando *t.getValue()* é o que retorna a palavra que foi escrita, sendo o único parâmetro necessário para utilizar na marcação de um ovo. Construiu-se então a função *procura_palavra()* que usa esse valor como parâmetro de entrada.

Essa função vai procurar dentro do *Array* das instâncias da classe *Bola*, na sua variável interna *tag* (que contém a palavra-pesquisa) se alguma destas corresponder à palavra que foi digitada na caixa de texto. Se essa comparação tiver um resultado verdadeiro, será criada uma nova instância de *Bola*, clone da instância que tem essa palavra-pesquisa, copiando todas as suas características. A única diferença desta instância é que este clone se situa sobre toda a amostra da visualização, sendo destacada através de uma cor alaranjada.

Esta caixa de texto não consegue ainda detectar palavras-pesquisa fora da amostra a ser visualizada, na medida em que apenas se estão a comparar as palavras do *Array* das instâncias de *Bola*, em vez de utilizar o ficheiro de entrada que contém a totalidade das pesquisas.

3.7 MÉTODOS E TÉCNICAS PARA A RECOLHA DE DADOS

A seguinte secção refere a recolha dos dados relativos da avaliação do protótipo realizou-se após a fase da sua implementação técnica. Procura avaliar de que maneira é que os utilizadores reagem e interagem com a aplicação, incidindo principalmente nos problemas encontrados e nas críticas construtivas referidas.

3.7.1 TÉCNICAS UTILIZADAS

No que concerne as técnicas de recolha de dados que foram seleccionadas, estas são a entrevista exploratória e o inquérito por entrevista.

A entrevista exploratória é realizada com o intuito de auxiliar a definição da problemática de investigação, enquanto ajuda a especificar o enquadramento teórico que se deve seguir. (Quivy & Campenhoudt, 1995) Esta foi efectuada com o Editor-chefe do portal Sapo, no início da investigação. Foi composta de perguntas abertas, onde o entrevistado poderia dar livremente sugestões e outras explicações sobre o seu trabalho diário. O seu objectivo seria aferir quais as principais carências do sistema de monitorização de dados pré-existente; foram também esclarecidos alguns pontos sobre a dinâmica do dia-a-dia da edição do portal, procurando-se discernir quais ferramentas de análise de dados mais relevantes neste contexto profissional. Foi a partir dos dados recolhidos desta entrevista que se formularam os objectivos da investigação.

Esta não será tida em conta na análise dos resultados, porque foi realizada com o objectivo de definir a pergunta de investigação e as respectivas finalidades e objectivos. Situa-se assim numa fase embrionária da investigação, sendo anterior à concepção do protótipo, não estando relacionada com a sua avaliação.

Cabe ao inquérito por entrevista este papel: será através deste que os dados de avaliação são registados e avaliados. Esta técnica pressupõe uma conversa entre o investigador e os participantes, onde estes

⁶⁰ Disponível em: <http://www.superstable.net/interfascia/>

partilham voluntariamente a informação de que dispõem. (Carmo & Ferreira, 1998) A construída entrevista é semi-dirigida, no sentido em que tanto é encaminhada por perguntas precisas, que envolvem uma escala numérica, como também inclui várias perguntas de resposta aberta. (Quivy & Campenhoudt, 1995, p. 192) Esta já foi realizada após a construção do protótipo funcional, com o objectivo de captar o contacto do público-alvo com a aplicação; procurou-se inquirir quais os pontos reconhecidos como mais importantes, assim como a identificação dos problemas principais.

3.7.2 CONSTRUÇÃO DOS INSTRUMENTOS

Na sua estrutura esta entrevista é dividida em 4 áreas distintas, que envolvem questões para análise quantitativa e qualitativa: Introdução, Representação Visual, Ferramentas de Interação e Perspectivas futuras. A parte da introdução recolhe dados de pessoais, tais como o nome, a idade e funções profissionais, assim como inquire sobre as expectativas que o utilizador poderá ter em relação a uma ferramenta deste género – o que é que lhe vem à cabeça sem conhecimento de causa que poderia existir nesta ferramenta.

A segunda parte aglomera perguntas sobre as características da metáfora adoptada, assim como da aplicação na generalidade.

O carácter interactivo da aplicação é tipo em conta na terceira parte da entrevista, onde são abordadas questões relativas às ferramentas de interação existentes.

A última parte efectua a conclusão da entrevista, focalizando-se nas perspectivas futuras que os utilizadores podem antever ou imaginar em relação ao potencial do protótipo, tendo em conta uma maior abrangência de dados representados, assim como aplicações em diferentes áreas de actividade.

Separaram-se as questões relativas à visualização e à interação, porque estas interpretam-se como sendo duas facções diferentes com regras diferentes. (Freitas, Luzzardi, Cava, Winckler, Pimenta, & Nede, 2002)

Os participantes justificaram usualmente as suas respostas verbalmente, como se pensassem alto (*think-aloud*) na resposta que pretendiam dar (tanto nas perguntas quantitativas como qualitativas).

A recolha dos dados foi realizada no dia de Maio, na Sede da Edição do Portal Sapo. Da parte da manhã foi realizada uma reunião, onde foi apresentado o protótipo, na presença de alguns agentes responsáveis pelo portal Sapo, incluindo o Editor-Chefe da redacção. Foi um momento informal onde estes deram o seu contributo verbal sobre a aplicação, e onde foram esclarecidos os pormenores acerca do momento de recolha de dados da investigação, a realizar nesse dia. Da parte da tarde, entre as 15h e as 18h, realizou-se essa recolha de dados, utilizando o próprio espaço da edição do portal, onde os inquiridos foram inquiridos à vez.

As entrevistas foram conduzidas após uma breve introdução ao contexto da investigação e às funcionalidades da aplicação, comunicados pelo entrevistador.

O enunciado escrito da entrevista foi sendo preenchido pelo entrevistador, à medida que os entrevistados iam respondendo às questões por ele efectuadas. Também se recorreu à gravação em áudio das entrevistas, através de um microfone ligado ao computador.

4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Após do momento de recolha de dados, procede a análise destes. De seguida são apresentados estes resultados e é efectuada uma análise crítica dos mesmos.

4.1 ANÁLISE QUANTITATIVA

Devido à dimensão reduzida da amostra de participantes, pode-se afirmar que não existe grande relevância nos resultados estatísticos da entrevista com o intuito de criar algumas generalizações. Fará então mais sentido tentar discernir quais os pontos que têm unanimidade de resultados e os que apresentam resultados mais díspares. Na resposta das perguntas de pergunta fechada verificou-se usualmente que os participantes justificavam verbalmente a sua escolha, mesmo sem terem sido solicitados. Este facto contribui para a análise dos resultados, pois permite encontrar mais facilmente justificações para as respostas mais contraditórias.

Não foram usadas medidas de tendência central (média, moda ou mediana), nem as correspondentes medidas de dispersão, pois existe pouca relevância estatística destes indicadores para estes resultados. Uma análise das percentagens relativas as diferentes respostas poderão ser mais interessantes, onde uma maior heterogeneidade de respostas merecerá uma análise mais cuidada.

As respostas foram divididas por área de investigação, nomeadamente a representação visual dos elementos e as ferramentas de interacção.

4.1.1 REPRESENTAÇÃO VISUAL

Nesta secção, pretende-se estudar as qualidades visuais da aplicação, e de que maneira é que a sua representação visual pode ser associada aos dados representados (pesquisas). No primeiro bloco de perguntas nesta categoria, foi pedido aos participantes que caracterizassem determinada característica visual da aplicação. A escala utilizada ia 1 a 4, onde 1 representa “Nada adequado” e 4 corresponde a “Muito adequado”.

a) RESOLUÇÃO DA APLICAÇÃO

Quanto à resolução da aplicação, 57% responderam adequada e 43% responderam muito adequada (Gráfico 1). Tal indica que não existem grandes objecções quanto a este factor, e tal confirmou-se pelo facto de não existirem grandes comentários verbais registados nesta pergunta. O facto de não existir uma taxa e 100% para muito adequada pode sugerir que a resolução pode ser optimizada, mas não há dados concretos que confirmem esta suposição.

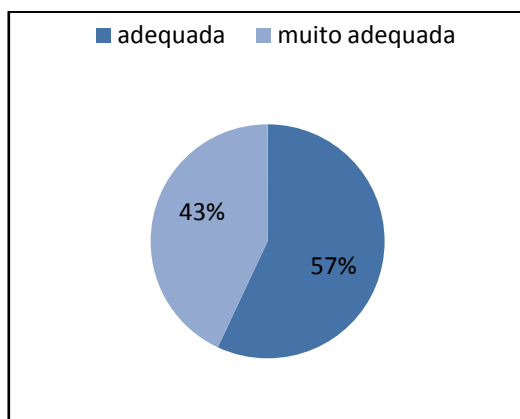


GRÁFICO 1 – RESOLUÇÃO DA APLICAÇÃO.

b) GAMA DE CORES

A gama de cores aplicada na aplicação originou a opinião unânime de 100% que esta é muito adequada (4). Tal factor indica que este factor também não deverá ser à partida um foco de objecções importante, e que a gama de cores poderá ter sido correctamente escolhida.

c) METÁFORA VISUAL (ESTÉTICA)

Quanto às opiniões que respeitam a qualidade estética da metáfora visual escolhida, da utilização dos ovos como elemento visual e da sua disponibilização no espaço, 14% consideraram esta como adequada e 86% como muito adequada (Gráfico 2). Mais uma vez não existiram objecções verbais em relação a este factor. Houve, no entanto, a necessidade de esclarecer os participantes de que esta pergunta apenas inqueria sobre o nível estético; verificou-se que as respostas estavam a ser dadas tendo em conta a funcionalidade da metáfora, que seria o objecto da próxima pergunta.

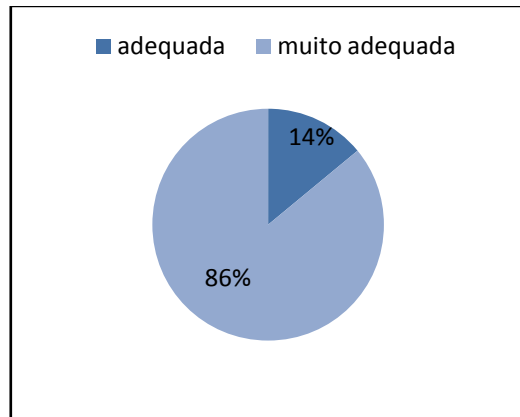


GRÁFICO 2 – METÁDORA VISUAL - ESTÉTICA.

d) METÁFORA VISUAL (REPRESENTATIVIDADE)

Quanto ao nível de representatividade da metáfora visual em relação aos dados (pesquisas efectuadas), as opiniões já se dividiram mais: 29% consideraram como pouco adequada, 42% consideraram como adequada e 29% consideraram como muito adequada (Gráfico 3).

A maioria transmitiu verbalmente a sua preocupação relativamente a eventuais problemas que poderão surgir ao analisar-se um maior conjunto de dados; por iniciativa própria, os participantes tenderam a considerar a falta de visibilidade de alguns elementos da visualização.

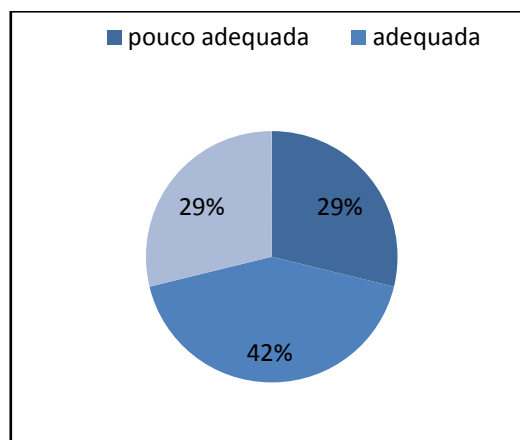


GRÁFICO 3 – METÁFORA VISUAL – REPRESENTATIVIDADE.

e) DENSIDADE (QT. DE ELEMENTOS / UNIDADE DE ESCALA)

A próxima pergunta utiliza uma escala de 1 a 5, onde 1 representa nada denso e 5 representa muito denso. A pergunta diz respeito à densidade da distribuição dos elementos dentro da aplicação. Esta foi considerada por 71% como densa e considerada adequada apenas por 29% dos participantes (Gráfico 4).

Em conjugação com os resultados da pergunta anterior, nota-se que existe uma grande preocupação dos participantes sobre a possível falta de visibilidade e legibilidade dos elementos da visualização, e que talvez a densidade seja um dos factores que contribui para este facto.

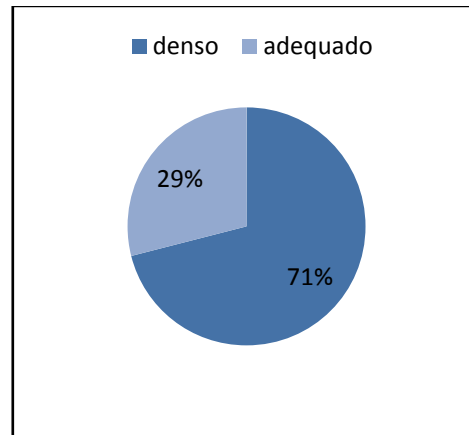


GRÁFICO 4 – DENSIDADE DA VISUALIZAÇÃO.

f) TAMANHO (GAMA DE TAMANHOS DOS ELEMENTOS DA VISUALIZAÇÃO)

Outro factor de visualização que pode ser importante para a compreensão da visualização é a distribuição do tamanho dos elementos que a constituem. Esta pergunta procura saber a opinião dos entrevistados acerca do tamanho, utilizando uma escala de 1 a 5, sendo 1 muito pequeno e 5 muito grande.

Os resultados foram que 14% dos inquiridos consideraram os tamanhos pequenos, 57% consideraram adequados e 29% consideraram grandes (Gráfico 5).

Neste caso não existe uma indicação tão clara de que o tamanho pode ser um problema; a maioria acaba por considerar este como adequado, e as opiniões variam entre pequeno e grande.

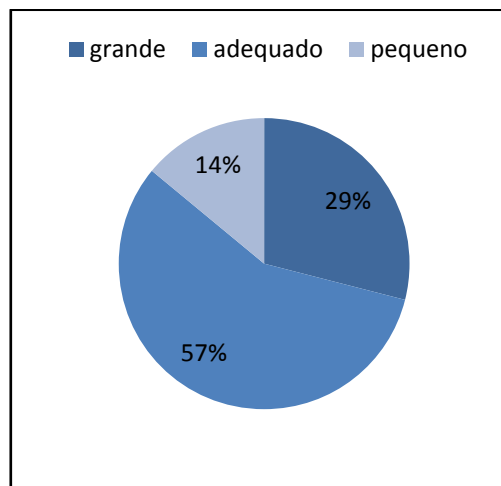


GRÁFICO 5 – TAMANHO DOS OVOS DA VISUALIZAÇÃO.

g) MODO DE VISUALIZAÇÃO DO CÍRCULO DE *POP-UP*

As perguntas seguintes focalizam-se mais na informação disponibilizada no círculo de *pop-up*, que surge com o clique directo num dos ovos. Esta pergunta procura saber em que medida é que é intuitiva a visualização desta informação. Para tal utiliza uma escala de 1 a 4, onde 1 representa “nada intuitivo” e 4 representa “muito intuitivo”.

Os resultados estão parcialmente divididos, mantendo uma maioria relativa: 57% dos inquiridos consideraram esta visualização como muito intuitiva, 14% consideraram-na intuitiva e 29% consideraram-na pouco intuitiva (Gráfico 6). Apesar da maioria das opiniões concordar que esta informação é muito intuitiva, ainda encontramos uma percentagem considerável quase no outro extremo. Existem algumas justificações anotadas durante a altura deste registo que podem ajudar a elucidar o porquê desta escolha. Uma das participantes afirmou que se não lhe dissessem que nunca adivinharia que os ovos seriam clicáveis, e portanto considerou que deveria haver algum tipo de notificação que alertasse essa opção. Outro participante considerou que, pela falta da indicação de qual o intervalo temporal de medição dos dados, existia alguma falta de clareza nos dados numéricos apresentados.

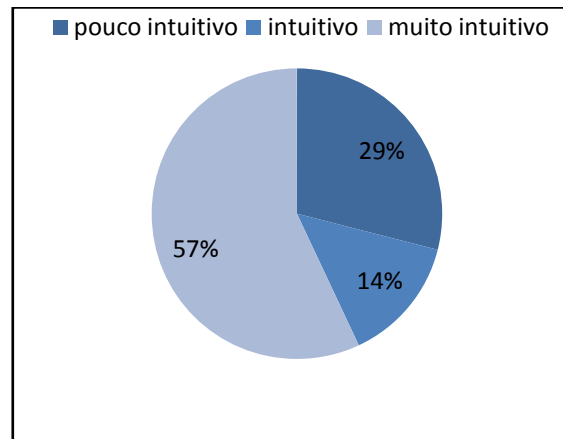


GRÁFICO 6 – MODO DE VER MAIS DETALHES.

h) RELEVÂNCIA DA INFORMAÇÃO DO CÍRCULO DE *POP-UP*

Esta pergunta, também relacionada com a informação inserida no círculo de *pop-up*, incide na relevância desta informação aos olhos do participante. A escala usada foi a mesma da pergunta anterior (1 a 5), com a diferença de que 1 corresponde a “nada relevante” e 5 corresponde a “muito relevante”. As respostas foram unânimes para esta pergunta: 100% dos inquiridos considerou que esta informação era “muito relevante”. A estas respostas seguiram-se comentários acerca da utilidade desta informação que é mais desenvolvida na parte 4, de perspectivas futuras, analisada na secção seguinte de análise qualitativa.

i) DIFICULDADE EM ENCONTRAR ELEMENTOS DE INTERACÇÃO

Esta pergunta procura aferir sobre o grau de dificuldade que os participantes consideram que tiveram para encontrar os elementos de interacção da aplicação. A escala disponibilizada é de 1 a 5, onde 1 representa “nada difícil” e 5 representa “muito difícil”. Na sua resposta, 14% dos inquiridos consideraram esta tarefa nada difícil, 57% consideraram pouco difícil e 29% consideraram adequado o grau de dificuldade (Gráfico 7). Tal pode ser indicador de que a mostragem das ferramentas de interacção é correctamente percebida.

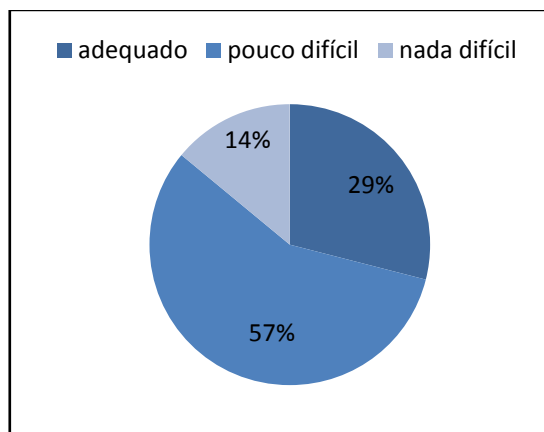


GRÁFICO 7 – DIFICULDADE EM ENCONTRAR ELEMENTOS DE INTERACÇÃO.

j) DIFICULDADE EM LOCALIZAR PALAVRAS-PESQUISA

Esta pergunta diz respeito à dificuldade que os utilizadores possam ter ao tentarem localizar visualmente palavras de pesquisa concretas no meio da visualização. A escala utilizada foi idêntica à pergunta anterior (1 a 5, 1 nada difícil, 5 muito difícil). Uma maioria dos intervenientes (58%), consideraram que esta tarefa é difícil, 14% consideraram o grau de dificuldade adequado, 14% consideraram pouco difícil esta tarefa, e 14% nada difícil (Gráfico 8).

A maioria dos participantes considerou a tarefa difícil, dando justificações verbais semelhantes, que apontavam o mesmo problema: o número de elementos na visualização adicionava confusão à leitura dos dados. Por outro lado, os participantes minoritários tiveram já em conta na sua resposta as ferramentas de interação apresentadas pela aplicação, nomeadamente, os *Sliders* de formatação da amostra, e a caixa de texto para pesquisa de palavras, que facilitariam essa tarefa.

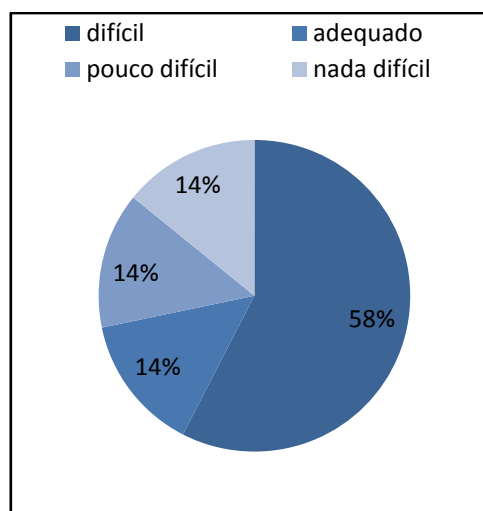


GRÁFICO 8 – DIFICULDADE EM ENCONTRAR DETERMINADAS PALAVRAS-PESQUISAS.

4.1.2 FERRAMENTAS DE INTERACÇÃO

Esta secção refere as interacções do utilizador com a aplicação, tendo em conta a funcionalidade com que este realiza algumas operações de manipulação da amostra de dados, e a importância que este atribui às ferramentas utilizadas.

a) FEEDBACK RECEBIDO NAS FERRAMENTAS DE INTERACÇÃO

A primeira pergunta desta secção procura classificar o nível de adequação dos estímulos visuais percebidos pelo utilizador, cada vez que altera estados através das ferramentas de interacção. Foi pedido aos participantes que dessem esta classificação através a escala já usada de 1 a 4, onde 1 representa “nada adequado” e 4 representa “muito adequado”. As notas são atribuídas especificamente para cada uma das seguintes ferramentas, que são ilustradas na Figura 42.

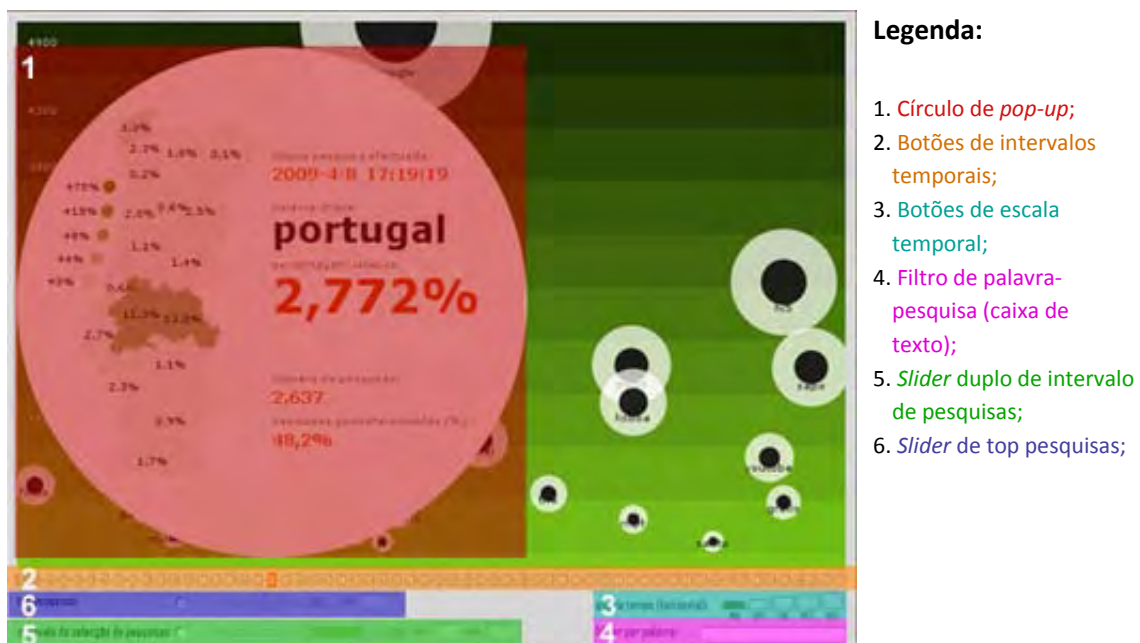


FIGURA 42 - DIFERENCIAÇÃO DAS DIFERENTES FERRAMENTAS DE INTERACÇÃO DA APLICAÇÃO.

Como se pode ver no Gráfico 9. As respostas dadas pelos participantes, no que diz respeito ao círculo de *pop-up*, foram: 57% consideraram o feedback visual muito adequado, 29% consideraram-no adequado e 14% consideraram-no pouco adequado; quanto aos botões dos intervalos temporais, 29% consideraram muito adequado, 14% consideraram adequado e 57% consideraram como pouco adequado; nos botões da escala temporal do eixo horizontal, 71% consideraram o seu feedback como muito adequado, e tanto o adequado como o pouco adequado foram considerados por 14% da amostra; relativamente à filtragem por palavra-pesquisa a partir da caixa de texto foi considerado por 86% da amostra como muito adequado e pelos restantes 14% como adequado; o feedback do *Slider* duplo que controla a selecção do intervalo das pesquisas foi encarado como muito adequado por 71% da amostra e os restantes 29% consideraram-no como adequado, obtendo-se os mesmos resultados para o *Slider* simples que controla o número do top das pesquisas.

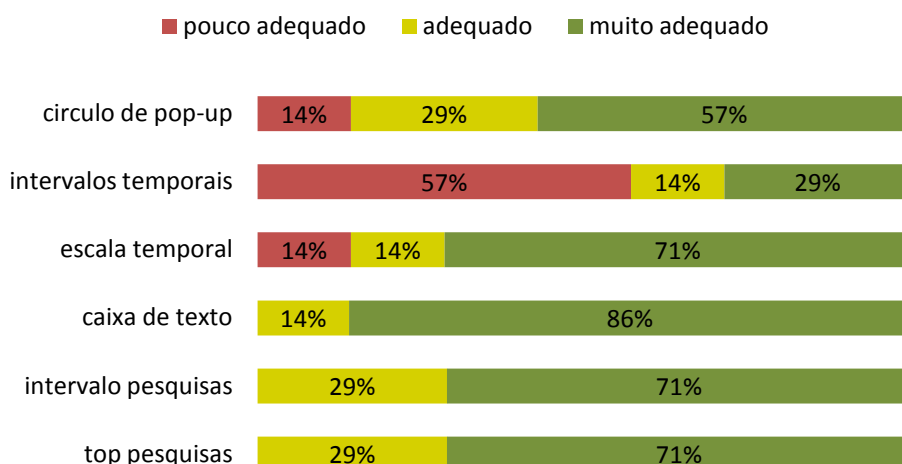


GRÁFICO 9 – FEEDBACK RECEBIDO QUANDO SE ALTERAM ESTADOS DA VISUALIZAÇÃO.

A partir destes resultados é possível verificar quais as ferramentas que sobressaem mais, tanto no bom como no mau sentido. As ferramentas referenciadas mais positivamente foram a caixa de texto da filtragem de palavra-pesquisa, e o *Slider* simples e duplo, relativos ao top pesquisas e intervalo de pesquisas, respectivamente. Quanto às ferramentas cuja utilização é mais difícil de compreender, os botões relativos aos intervalos temporais foram os mais criticados, tanto por estes resultados apresentados como verbalmente, pelos participantes. As principais dificuldades referidas foram tanto relativas à compreensão da ferramenta como interactiva, como à compreensão do que é que esta alterava na visualização.

b) IMPORTÂNCIA DAS FERRAMENTAS DE INTERACÇÃO

A mesma estrutura e escala da pergunta anterior (pergunta 2.12 da entrevista – ver Anexo 1) foram adoptadas para a seguinte. Ao invés de inquirirmos sobre o feedback visual, procura-se avaliar como os participantes consideravam a importância de cada uma das mesmas ferramentas na aplicação. A escala manteve-se de 1 a 4, com a diferença de que 1 representa “nada importante” e 4 representa “muito importante”.

Os resultados para esta pergunta são um pouco mais positivos. O círculo de *pop-up* é reconhecido como muito importante por 86% dos participantes e importante por 14% destes; os botões relativos aos intervalos temporais são considerados por 57% como muito importantes, importantes por 29%, e pouco importantes por 14% da amostra; os botões da escala temporal são reconhecidos como muito importantes por 57% e importante por 43% dos entrevistados; a caixa de texto é unanimemente considerada como muito importante; o *Slider* duplo do intervalo de pesquisas foi considerado como muito importante por 43%, e importante por 57% dos participantes; o top de pesquisas é considerado como muito importante por 71%, e importante por 29% da amostra (Gráfico 10).

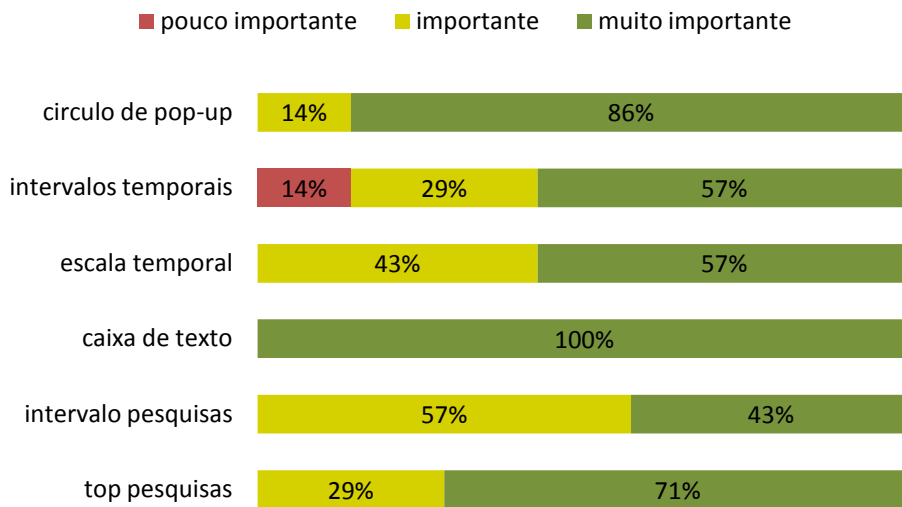


GRÁFICO 10 – GRAU DE IMPORTÂNCIA DAS FERRAMENTAS DE INTERACÇÃO

Através da análise destes resultados é possível à partida afirmar que todas as ferramentas são consideravelmente importantes; a ferramenta da filtragem de palavras-pesquisa específicas parece ser à partida a mais importante da listagem. O círculo de *pop-up* que surge ao clicar num ovo também é reconhecido como muito importante pela grande maioria dos entrevistados.

Uma comparação directa entre as ferramentas do top pesquisas e do intervalo de pesquisas mostra uma certa vantagem na ordem de importância para a primeira. Ambas as ferramentas têm funções semelhantes, na medida em que alteram a amostra de dados, mas o top de pesquisas parece ser a mais relevante. Também a escala temporal é uma das mais importantes e ser considerada. Por outro lado os botões relativos aos intervalos temporais obtêm a única percentagem de estes serem poucos importantes, apesar de manter a mesma percentagem que a escala temporal relativa a muito importante. Pode-se dar o caso de que esta percentagem de pouco importante (14%) esteja associada à falta de compreensão, registada na pergunta anterior. Por ser pouco adequada, também pode significar que seja menos importante.

4.2 ANÁLISE QUALITATIVA

Uma outra parte importante para a avaliação do protótipo reside na sua parte respectiva à análise qualitativa; a maioria das opiniões e problemas mais completos foram levantados pelos participantes foram transmitidos verbalmente, através de um discurso informal, tal como um *think-aloud*, onde os participantes vão explorando a aplicação e ao mesmo tempo falando com o entrevistador sobre os problemas que vão encontrando. Neste caso passou-se algo semelhante, onde os participantes iam justificando as suas perguntas (mesmo quando não solicitados).

Apesar da grande relevância destes dados, a amostra de estudo continua a ter uma dimensão demasiado reduzida para que se necessite de uma análise de conteúdo aprofundada.

Considerou-se então mais exequível para esta investigação, a análise destes dados através de uma descrição detalhada dos pontos mencionados pelos participantes, incidindo-se com maior destaque nos que foram mais referidos. As opiniões isoladas, apesar de não serem uma opinião corrente, são descritas se forem consideradas relevantes para a investigação.

4.2.1 EXPECTATIVAS

A primeira pergunta descritiva (questão 1.2, ver Anexo 1) diz respeito às expectativas que os utilizadores, nomeadamente a equipa do Portal Sapo, possam ter sobre a aplicação. Na sua análise dos dados, verificou-se a intenção generalizada dos participantes em obterem respostas sobre os temas de maior procura e interesse das pessoas, tanto utilizadores do portal Sapo, como num contexto mais generalizado de utilizadores. A distribuição destas pesquisas por região surgiu também como um indicador interessante.

Foi transmitido também um interesse especial dos participantes em visualizar dados das pesquisas de notícias realizadas no portal, e o intuito principal de se perceber quais as notícias mais pesquisadas. Este interesse revelou-se tanto numa perspectiva de análise a curto prazo como a longo prazo. Nas análises de curto prazo interessariam mais as notícias pesquisadas mais recentemente e a sua frequência; a longo prazo, seriam interessantes a possibilidade de análise das várias notícias pesquisadas, agrupadas por grupos temáticos de pesquisa definidos pelo utilizador. Também a construção de gráficos comparativos ao longo de um ano se revelou como uma ferramenta de interesse, a fim de se poderem detectar quais os tópicos mais pesquisados por alturas diferentes do ano.

Quanto à Georreferenciação dos dados, apenas um sujeito mostrou-se interessado em saber de onde é que eram realizadas as pesquisas, reconhecendo que seria uma ferramenta mais relacionada com a sua área (o *Sapo local*).

4.2.2 REPRESENTAÇÃO VISUAL

Relativamente a sugestões enunciadas face à visualização dos dados disponibilizada na aplicação, tanto no círculo de *pop-up* disponibilizado para ver mais detalhes, como na visualização da amostra no geral (pergunta 2.9, Anexo 1), destaca-se:

- a) Um interesse especial em tornar visível o período de tempo da amostra recolhida, de quando a quando é que foi medida, em vez de ser apenas mostrada a data da última pesquisa efectuada.
- b) A sugestão de se dar mais relevância ao número absoluto de pesquisas, recolhido numa janela temporal, do que à percentagem relativa de pesquisas, considerando o primeiro indicador mais importante.
- c) No que diz respeito à disponibilização dos ovos no eixo bidimensional, também foi referida a importância de estarem presentes legendas para cada um dos eixos de forma mais explícita, que indiquem o que cada um representa em concreto.
- d) A indicação do número de pesquisas por defeito em cima de cada ovo, também foi mencionada como um possível melhoramento.
- e) Quando se dá o alinhamento horizontal dos ovos, numa unidade temporal, é considerado necessário encontrar formas de distinguir perceptivamente os ovos que ficam sobrepostos.

4.2.3 FERRAMENTAS DE INTERACÇÃO

Quanto a ferramentas de interacção adicionais que poderiam ser adicionadas à aplicação (pergunta 3.2, Anexo 1), surgiram várias observações:

- a) Foi sugerido que na escala temporal, quando esta é disponibilizada, que fosse possível clicar em qualquer um dos elementos a qualquer altura, assim como poder clicar no eixo

secundário temporal, que surge a após a selecção de determinado ciclo de tempo (uma hora desdobra-se em minutos, um dia desdobra-se em horas, por exemplo).

- b) Ao clicar no ovo, também se poderia mostrar uma listagem de notícias do portal relacionadas com essa palavra, ou a eventual listagem dos resultados da pesquisa que se obteriam no motor de busca para essa palavra.
- c) A pesquisa de várias palavras em simultâneo e a comparação directa dos seus dados foi apontada como interessante.
- d) Tal como referido nas expectativas dos participantes (página 80), a análise a longo prazo também foi uma opção a ser muito referida, e a possibilidade de registar dados personalizados para posterior comparação de dias num espaço de tempo alargado.

Sobre as ferramentas de interacção já criadas também foram sugeridas alterações:

- a) o surgimento de um *pop-up* de ajuda que explicasse rapidamente a função de cada uma das ferramentas.
- b) Um mecanismo de *rollover* também poderia estar activo nos ovos, de maneira a que a bola sobre a qual se passasse o rato se sobrepusesse às outras, de forma a destaca-la rapidamente das restantes.

Quanto a mecanismos alternativos de interacção (pergunta 3.3, Anexo 1), em alternativa ao teclado e rato, esta medida foi considerada irreal e dissociada dos hábitos de trabalho dos participantes. Houve quem considerasse a tarefa de arrastar o rato como penosa: no caso dos Sliders deveria ser possibilitado o clique directo numa zona dentro dos limites desta ferramenta, que causasse o deslocamento automático do *Slider* para a posição clicada.

No entanto, algumas alternativas para o rato e teclado chegaram a ser mencionadas: o *touch screen* e um comando do tipo *Wii*⁶¹, mas sempre considerando um cenário utópico ou futurista, fora da realidade profissional.

Na questão sobre a visualização imediata versus a exploração interactiva (pergunta 3.4, Anexo 1), as opiniões dividem-se, mas na generalidade é reconhecida a importância das duas modalidades na análise de dados, reconhecendo que cada um será específico para objectivos diferentes. A interactividade é reconhecida como sendo mais apelativa, mas a função já existente nos monitores da edição Sapo também é considerada importante; a primeira será mais útil para análises a longo prazo, ao longo do tempo e com um conjunto alargado de dados; a segunda para uma análise mais a curto prazo, para suportar decisões mais imediatas que baseiam-se na reacção do público no momento.

Outra hipótese considerada interessante consiste na faculdade do utilizador em escolher quais os dados a visualizar constantemente no ecrã, sempre que quisesse observar apenas alguns indicadores concretos, não estando limitado aos dados que lhes eram mostrados por defeito.

4.2.4 PERSPECTIVAS FUTURAS

De seguida os participantes foram inquiridos sobre que outros dados seriam relevantes para uma representação visual tendo em conta o portal Sapo (pergunta 4.1, Anexo 1). Uma opinião recorrente verificada foi a de que seria mais interessante e útil a utilização dos dados de pesquisa das notícias do portal, ao invés das pesquisas gerais. De modo complementar, também surgiu a ideia de fornecer a listagem de notícias do portal que estejam relacionadas com a palavra-pesquisa. Alguns participantes consideraram

⁶¹ Consola de jogos da marca Nintendotm, que usa um comando *wireless* que dá a ideia de interacção directa com o ecrã.

relevante a generalização desta abordagem de visualização para dados mais alargados, tais como a listagem de *blogs* e fotos mais acedidas.

Também foi sugerida a disponibilização de ferramentas de análise das saídas do portal para determinados domínios externos, através da leitura dos URLs⁶².

A pergunta final da entrevista diz respeito à viabilidade da utilização de uma aplicação deste género em outras áreas de actividade (pergunta 4.2, Anexo 2). Nas suas respostas, foi recolhida a opinião unânime de que esta aplicação poderá ter várias utilizações interessantes em várias áreas. Entre as áreas de actividade referidas, contam-se quaisquer cadeias comerciais que necessitem de saber o que as pessoas procuram. A análise de informação deste género poderá contribuir na decisão do lançamento de determinados produtos, assim como aferir sobre qual o impacto desse lançamento nas procuras do consumidor. Neste ponto de vista, a Georreferenciação das pesquisas poderá ser útil também para tomadas de decisões que envolvam medidas mais apropriadas consoante os locais geográficos, a partir da análise das zonas com maior e menor interesse em determinados tópicos (produtos e serviços). A informação recolhida destas análises poderá auxiliar a criação de estratégias de *marketing* mais personalizadas, que tenham em conta as particularidades regionais.

⁶² URL – *Uniform Resource Locator*, usualmente conhecido como o endereço de uma página *web*.

5. CONSIDERAÇÕES FINAIS

Nesta secção reflectem-se os resultados obtidos na avaliação do protótipo, com o intuito de encontrar pontos que devem ser melhorados e de que forma é que este melhoramento se pode dar, para que este se adapte mais às necessidades do público-alvo. Também é considerado todo o processo de concepção e implementação deste protótipo, no que respeita os problemas encontrados e quais as possíveis prevenções que se possam efectuar para futuras investigações.

Por fim, analisar-se-á esta investigação de um ponto de vista mais holístico, e pensar que novos estudos e trabalhos poderão advir deste, que possam colmatar as fragilidades do protótipo construído, ou elaborar novas propostas de visualizações.

5.1 ANÁLISE DA PERGUNTA DE INVESTIGAÇÃO

Interessa, nesta fase final da investigação, analisar a pergunta e as hipóteses (ver na secção 1.1), formuladas inicialmente, de forma a compreender como é que estas foram exploradas, e apreciar em que medida é que estas se complementaram, considerando todo o processo de criação do protótipo.

Ambas as hipóteses da investigação procuravam responder à pergunta de investigação e foram, de certa forma, aplicadas funcionalmente no protótipo. Não obstante, deram-se algumas alterações formuladas ao longo da investigação, inerentes à própria natureza iterativa do processo de conceptualização e implementação de uma aplicação multimédia. Muitas destas alterações surgiram em resposta a alguns problemas encontrados (referidos na secção 5.2), o que levaram à mudanças de direcção das hipóteses propostas inicialmente.

No que respeita à pergunta de investigação convém relembrá-la como:

Que proposta de metáfora visual poderá ser conceptualizada e implementada com base nos dados do broker resultantes das pesquisas realizadas no portal Sapo?

Foram desenvolvidos vários passos ao longo da investigação que procuraram chegar a uma possível resposta para esta pergunta.

Uma entrevista exploratória foi conduzida com o Editor-chefe do portal Sapo, onde se aferiram os interesses do entrevistado em relação aos dados disponibilizados pelo sistema Sapo *Broker*.

Após a análise das respostas recolhidas, elaborou-se um conjunto de hipóteses e objectivos, orientadores para a investigação.

Através destes objectivos procedeu-se à análise bibliográfica dos diferentes tópicos teóricos considerados relevantes para o seu desenvolvimento. Esta pesquisa bibliográfica tanto se deu a um nível teórico, através da exploração da área temática da Visualização de Informação, como a um nível prático, em que se desenvolveram os conhecimentos técnicos necessários para a elaboração do protótipo.

Com base nos conhecimentos adquiridos, procedeu-se à conceptualização do protótipo. Esta começou com a análise dos tipos de dados, disponibilizados pelo sistema Sapo *Broker*, relacionados com a questão de investigação (dados de pesquisas). Com base nestes tipos de dados estudaram-se que diferentes tipologias de visualização poderiam ser as mais adequadas para a sua representação, sob a forma de metáforas visuais, assim como as ferramentas de interacção mais relevantes para a manipulação dos dados (ver capítulo 3.5, página 44).

Simultaneamente, também foram estudados os procedimentos técnicos utilizados na recolha de dados do Sistema *Broker*, com o intuito de proceder à recolha de uma amostra destes dados. Esta amostra seria posteriormente utilizada para o desenvolvimento do protótipo.

A implementação do protótipo deu-se em ambiente *Processing*. Esta ferramenta foi seleccionada considerando que é bastante utilizada em projectos do género (Visualização de Informação), e tem uma livreria extensa *open source*. A fase da implementação foi executada iterativamente, com a adição ou subtracção de certas componentes (visuais e interactivas), até se chegar a um protótipo funcional considerado fiel aos objectivos elaborados.

Por fim, deu-se o momento de avaliação da aplicação, através da inquirição de uma amostra do público-alvo definido – a equipa de edição do Portal Sapo. Através da análise das respostas adquiridas procurou-se encontrar problemas no *design* da aplicação, assim como encontrar novas sugestões para futuras remodelações da mesma.

Foram formuladas duas hipóteses que procuravam descrever respostas possíveis para a pergunta de investigação, e que se tornaram as directrizes principais para a construção do protótipo. Estas hipóteses consideram duas tipologias diferentes de implementação.

Relembra-se a primeira hipótese formulada, como resposta à pergunta de investigação:

Através da representação de uma amostra de dados, que permita de um modo rápido verificar picos de pesquisa em tempo real, dando pouco ênfase às suas características particulares. Cada dado de pesquisa poderá ser representado por um corpo individual, cuja diferenciação dos outros corpos poderá dar-se a um nível morfológico e/ou espacial.

Esta hipótese sugere um sistema de uma amostra passiva de dados, representados através de figuras metafóricas. Estas figuras serão diferenciadas entre si através de morfologias ou movimentações diferentes, consoante os valores que representam.

Na construção do protótipo, esta metáfora foi operacionalizada através da já apresentada visualização das pesquisas através de ovos, que utilizam regras semelhantes de flutuação, mas com tamanhos e alturas diferentes (explicado na página 57, Secção a)).

No que concerne a componente passiva da visualização sugerida na hipótese, esta foi posta de parte aquando da entrevista exploratória. O entrevistado mostrou um interesse especial numa aplicação interactiva, ao invés de uma aplicação passiva, o que levou à formulação de objectivos da investigação que incidiam na manipulação interactiva do utilizador. Consequentemente foram introduzidas ferramentas de interacção que permitem ao utilizador controlar a direcção da sua análise dos dados, manipulando a amostra que estava a ver.

Esta decisão foi apoiada na fase posterior da avaliação do protótipo, onde os participantes reconheceram, de um modo geral, a importância da interactividade do sistema, defendendo, ao mesmo tempo, a manutenção de uma solução passiva de visualização, na mesma útil para propósitos diferentes de análise.

Quanto à segunda hipótese formulada, esta consiste em:

A distribuição geográfica das pesquisas efectuadas poderá ser representada numa janela temporal, que transmita imediatamente quais os distritos do país com uma maior afluência de pesquisas, sendo possível ao utilizador explorar os dados, permitindo-lhe ver um nível personalizado de detalhes.

Esta hipótese incide mais nos dados georreferenciados, e surgiu numa primeira instância associada à Geovisualização pré-disponível na Edição Sapo, como uma elaboração do mesmo, onde se pretendia mudar a sua aparência e adicionar ferramentas interactivas.

Com o decorrer da investigação, esta solução foi-se distanciando gradualmente do modelo conceptual da primeira hipótese, e começaram então a surgir dúvidas relativas ao nível de compatibilidade entre as duas visualizações resultantes - os ovos de pesquisas (Secção 3.5.1, página 45) e o mapa de pesquisas (Secção 3.5.2, página 48). Tal incompatibilidade surgia principalmente ao nível dos objectivos na análise dos dados; a primeira visualização incidiria mais na análise de palavras-pesquisa individuais, e na sua relação com as outras; a segunda reflectiria um panorama geral das pesquisas, sem diferenciação das palavras pesquisadas.

Face a esta dificuldade, e com o intuito de chegar a um protótipo visualmente coeso, a segunda hipótese foi operacionalizada de forma diferente. A Georreferenciação passou a ser específica a cada palavra-pesquisa: o mapa do país relativo às estatísticas das pesquisas totais efectuadas passou a representar a distribuição geográfica das ocorrências para cada palavra distinta (Figura 27, página 50).

Esta nova abordagem conseguiu atenuar as diferenças entre as duas hipóteses, e foi na generalidade bem recebida pelo público-alvo, que considerou a informação importante assim como o seu modo de visualização fácil de perceber.

5.2 LIMITAÇÕES DO ESTUDO

Antes de se avaliarem os resultados finais, é importante ter em conta as várias dificuldades e limitações que foram encontradas no decorrer da investigação. Mais especificamente, interessa dar ênfase àquelas que poderão ter influenciado mais directamente o processo de implementação e avaliação do protótipo.

5.2.1 LIMITAÇÕES NA IMPLEMENTAÇÃO

Quanto à própria implementação técnica do protótipo, surgiram alguns contratempos que limitaram algumas escolhas, relacionados em concreto com algumas das funcionalidades inicialmente propostas:

- a) Um objectivo inicialmente ponderado não posto em prática nesta investigação. Este diz respeito à análise da reincidência de determinados utilizadores na pesquisa de certos tópicos. Este objectivo consistiria, mais especificamente, no acompanhamento progressivo de determinados *IPs* na pesquisa de tópicos pré-estabelecidos, ao longo do tempo. Pretender-se-ia com isto tentar perceber, na rotina diária dos utilizadores, quais as alturas do dia em que determinados tópicos de pesquisa seriam mais procurados.

Durante algumas sessões informais com alguns responsáveis do portal Sapo que foram acompanhando o trabalho desenvolvido, fomos recebendo algumas indicações e esclarecimentos. Numa destas sessões, foi sugerido o abandono desta ideia de implementação, sendo então explicada a existência de problemas legais relacionados com o seguimento de *IPs* específicos dos utilizadores. Acresce ainda que estes dados são à partida filtrados antes de serem consumidos, pelo que não teríamos acesso a esta informação.

- b) Outro objectivo proposto para o projecto seria a representação em tempo real (ou durante intervalos reduzidos de tempo, medidos continuamente) dos dados de pesquisa, permitindo ver assim de forma fluida a alteração dos picos de pesquisa, através de mudanças da movimentação vertical dos ovos. Este objectivo não foi desenrolado por causa do próprio processo de leitura e gravação dos dados do *Brooker*.

Em vez de uma representação em tempo real, foi criado um protótipo que representasse os dados retirados apenas numa janela temporal definida manualmente pelo investigador. Esta tipologia foi escolhida porque se verificou na altura que o processo da leitura de dados não podia ser acedido de forma contínua e ilimitada, como seria ideal, para a testagem do protótipo. A aquisição dos dados só foi realizada em alguns momentos da investigação, com a ajuda de um docente da Universidade com acesso a esta ferramenta, sendo a utilização destes dados feita só posteriormente a esta recolha.

- c) A limitação referida anteriormente (alínea b) também influenciou o cancelamento de outra funcionalidade considerada inicialmente para este protótipo: permitir ao utilizador a manipulação directa do parâmetro temporal da análise dos dados; ou seja, permitir a escolha do período de tempo durante a qual o sistema efectuar a recolha e análise dos dados.

- d) A análise a longo prazo dos dados não foi considerada para esta investigação. Caso a janela temporal da recolha de dados tivesse mais do que algumas horas, a quantidade de dados seria astronómica, demasiado grande para ser armazenada em ficheiros de texto. Seria então necessário existir um sistema de gestão de base de dados, que utilizasse, de preferência, um servidor da Sapo. Este SGBD⁶³ filtraria, já no próprio processo de consumo de dados, apenas os dados relevantes, poupando recursos de *hardware*.

- e) Outro contratempo encontrado afectou a precisão estatística dos dados, mais especificamente na Georreferenciação das pesquisas. Como referido anteriormente, o cálculo da região do país para cada pesquisa efectuada não foi o mais adequado e completo, podendo apenas servir como uma solução temporária. Este processo, consistindo na determinação da região através da cidade de pesquisa, só

⁶³ SGBD – Sistema de Gestão de Bases de Dados

funciona para pesquisas efectuadas em cidades-concelho. As pesquisas realizadas noutras localidades serão geograficamente desprezadas. Esta medida foi adoptada porque não se conseguiu averiguar como é que se dava a Georreferenciação no sistema já existente na redacção Sapo.

Foram trocados alguns *e-mails* com pessoas integradas nos dados de pesquisa do *Brooker*, para aferir sobre esta situação. Em resposta, foi-nos então confirmada a dúvida que já existia sobre a impraticabilidade da *tag <region>* para este cálculo, que no momento não estaria a usar nenhuma escala numérica representativa das regiões geográficas. No sistema pré-existente, estaria a ser efectuado outro tipo de cálculos para determinar a região, mas não nos foi acedida informação sobre esse método. Foi então necessário recorrer a outra maneira provisória de recolher estes dados; a primeira opção a ser considerada foi a utilização dos dados encontrados nas *tags <longitude>* e *<latitude>*, presentes no pacote de dados das pesquisas. Esta solução provou-se com o tempo mais complexa do que o esperado, e teve de ser abandonada devido a pressão do horário. Por fim foi adoptada a solução explicada anteriormente (na secção do *Design técnico*, página 55), da comparação de cidades-concelho - apesar de representar uma Georreferenciação menos completa, continua a ser adequada para os propósitos deste protótipo.

5.2.2 LIMITAÇÕES NO PROCESSO DE AVALIAÇÃO

No que diz respeito ao processo de recolha de dados que foi utilizado para a avaliação do protótipo, este também não decorreu conforme todas as expectativas.

Durante a própria formulação da ferramenta de recolha de dados, não nos foi possível integrar nesta mais perguntas que permitissem um estudo aprofundado de usabilidade da aplicação; também teria sido interessante realizar um enunciado de tarefas, que pedisse ao utilizador a realização de simples acções de interacção, com o objectivo de observar determinados resultados na visualização. A execução deste resultado teria optimizado a veracidade das perguntas realizadas posteriormente, no inquérito por entrevista.

Estas opções não foram tomadas porque exigiriam um maior dispêndio de tempo dos inquiridos, o que não seria razoável; isto tendo em conta que a recolha de dados foi realizada durante o seu horário de trabalho, e não seria indicado privá-los deste horário durante muito tempo.

Ao invés disso, coube ao entrevistador apresentar sumariamente o protótipo no início da sessão: explicar qual o significado da visualização e dos dados que esta representa, assim como a apresentação das ferramentas de interacção, e o que estas realizam no sistema.

Desta maneira, é possível que o entrevistador tenha influenciado o entrevistado na sua percepção do protótipo, visto que já o estaria a ajudar na sua utilização. Felizmente, a maioria dos entrevistados teceram críticas construtivas nesta questão, reconhecendo esta possibilidade, e procurou dar então a sua opinião com base na sua primeira impressão, muitas vezes verbalizando: “se não me tivesse dito que isto fazia tal, não conseguiria perceber isso”, por exemplo.

O tempo disponibilizado para realizar a totalidade da recolha de dados também foi limitado (cerca de 3h), o que não permitiu que muitos participantes fossem inquiridos, considerando que cada sessão delongou-se de 15 a 20 minutos. Tendo em conta estes constrangimentos temporais, apenas se conseguiu entrevistar uma amostra de 7 pessoas, cuja relevância estatística para chegar a conclusões generalizadoras é impossível.

Já durante o próprio decorrer da recolha de dados, existiram participantes que, ou por falarem demasiado baixo, ou acidentalmente desligarem o microfone durante a entrevista, ficaram com registos sonoros imperceptíveis ou incompletos. Isto implica que algumas respostas tenham ficado apenas registadas através das anotações do entrevistador, sendo possível que alguma informação tenha ficado incompleta.

Para prevenir a situação decorrida, teria sido preferível que as entrevistas se tivessem dado numa sala privada, com pouco ruído de fundo, e que fosse utilizado um gravador áudio para a sua gravação, sem que os entrevistadores tivessem de ter algum contacto com este. Tais condições não se proporcionaram para esta investigação.

5.3 ANÁLISE DO PROTÓTIPO DESENVOLVIDO

Tendo em conta o *feedback* recebido pelos participantes inquiridos - especialmente o que foi recolhido através da análise de conteúdo - podemos identificar à partida quais as questões mais relevantes que poderiam ser revistas no protótipo implementado. Tal como em secções anteriores, é relevante dividir estas questões nas duas categorias principais que compõem o protótipo: Visualização e Interacção.

a) VISUALIZAÇÃO

A falta de uma correcta legendagem dos eixos, assim como da disponibilização de outra informação elementar (tal como o período de tempo da recolha da amostra) foram questões muito mencionadas pelos participantes.

Informação deste tipo poderá ser muito importante para que o utilizador consiga ter uma contextualização do que está a ver. Esta contextualização foi uma das componentes que acabou por não ser muito explorada no protótipo, porque foi dada primazia ao factor estético da visualização, que passa pela ideia de minimalismo. No entanto, este minimalismo não se adequa, a partir de certo ponto, a uma Visualização de Informação, especialmente se esta pretender ser uma ferramenta completa de análise de dados. Será, a certa altura, necessário o apoio textual na visualização, o que inclui as devidas medidas de escala e legendagens, para permitir uma análise ponderada e mais aprofundada.

Da mesma forma, não foram implementados nenhuns mecanismos de ajuda ao utilizador, o que também foi referido pelos participantes. Estes apontaram sobretudo a falta de informação disponibilizada sobre quais as utilidades das ferramentas de interacção do sistema, e quais as suas funções. Esta funcionalidade será mais relevante numa implementação posterior à fase de protótipo, após estar definida toda a estrutura principal do programa.

No que diz respeito à navegação do utilizador dentro da visualização, um problema comum de especial destaque é o da sobreposição dos ovos, e a sua consequente ilegibilidade individual; as palavras-pesquisa com valores semelhantes de ocorrências podem surgir em posições semelhantes o que origina a que se sobreponham, sendo apenas um dos ovos clicável, em detrimento de outros que ficam inalcançáveis. Com base neste facto, revela-se a necessidade de pensar em estratégias (de visualização ou de interacção) que procurem atenuar esta dificuldade de navegação.

Uma destas possíveis estratégias que podem ser adoptadas, sugeridas até por um dos participantes, poderá passar pelo uso do evento de *rollover* do rato, sobre os elementos em sobreposição. Este evento, para além de ser razoavelmente simples de implementar não implica à partida uma acção directa do utilizador (tirando a movimentação do rato por cima dos elementos); poderá ser interessante na medida em que podem surgir muitas acções possíveis deste acto. Uma das acções que este poderia executar seria a listagem ou afastamento espacial dos elementos sobrepostos, de forma a possibilitar ao utilizador ver simultaneamente as informações relativas todos estes elementos. Tal seria interessante para o caso de se pretenderem comparar palavras diferentes com números de pesquisas semelhantes.

Existe no entanto, todo um número indeterminado de soluções que deveriam ser estudadas para resolver este pormenor.

Quanto aos próprios valores indicadores da amostra de dados, verificou-se que os participantes dão mais importância aos valores absolutos do que a percentagens relativas. Isto implica que talvez não seja preciso esta transposição (de absolutos a relativos). Tendo em conta a experiência do utilizador, é possível que este já esteja de tal forma familiarizado com a gama de valores absolutos (número de pesquisas) que as pesquisas podem ter, que já consiga só através deste valor detectar focos que mereçam interesse. Se tal fosse verdade, poder-se-ia simplificar consideravelmente os parâmetros da visualização: o tamanho dos ovos já não precisaria de representar uma percentagem relativa à amostra, mas podia ser utilizado para representar outro indicador, tal como a frequência de ocorrências na última janela temporal.

Quanto à possível relação entre a estética da visualização e a sua funcionalidade, não é possível concluir uma relação directa entre as duas, visto que as opiniões divergiram, sendo a estética reconhecida quase unanimemente como muito adequada, enquanto que a representatividade causou mais diferenças de opiniões.

b) INTERACÇÃO

Já nas questões relacionadas com as ferramentas de interacção, tendo em conta as reacções dos utilizadores, uma ferramenta específica revelou-se mais problemática: a ferramenta que lida com a selecção de intervalos de tempo específicos (Figura 40, página 66).

O modo de aceder a cada tempo de pesquisa, nas suas diferentes dimensões (ano, mês, dia, hora, minuto) foi pouco compreendido pelos participantes, assim como o modo de sair da ferramenta (o modo de desactiva-la). Apesar da principal causa para esta dificuldade residir no *design* defeituoso da ferramenta, também acresce a dificuldade de utilizar a mesma metáfora visual para objectivos de análise diferentes. Para este protótipo utilizou-se a mesma visualização para incorporar, tanto a representação de uma amostra de dados em constante mudança (em tempo real simulado), como uma disposição ordenada por tempo. Tal facto levanta questões sobre a possível relevância de se utilizarem diferentes esquemas de visualização consoante o tipo de análise, seja esta a curto ou a longo prazo.

Nas considerações respeitantes a formas de interacção alternativas, para além do rato e do teclado, estas foram encaradas com pouco interesse e algum cepticismo em relação à sua aplicabilidade. Os participantes consideraram mais adequado o uso dos mecanismos convencionais de interacção, baseando-se não só sua experiência profissional, mas também nos seus hábitos e ferramentas de trabalho actuais.

Finalmente, será relevante analisar a preferência dos participantes entre um sistema de visualização em tempo real, que transmita informação de uma forma passiva mas imediata, e uma aplicação interactiva que permita uma análise baseada na exploração dos dados. De um ponto de vista geral, os participantes consideraram que as duas vertentes eram úteis, com objectivos de uso diferentes, mas complementares; a primeira solução, à semelhança da já existente na Edição Sapo, seria utilizada para uma ajudar a tomada de decisões a curto prazo, interessando apenas transmitir ao utilizador a visão geral das pesquisas mais relevantes do momento; a segunda vertente já seria interessante para se realizarem análises detalhadas sobre a evolução a longo prazo das pesquisas efectuadas, de um modo colectivo ou particular.

5.4 PERSPECTIVAS DE TRABALHO FUTURO

Tendo em conta a área profissional do público-alvo, não foi de todo inesperada a prevalência de um interesse generalizado, da parte dos participantes, em replicar a solução desenvolvida a novos contextos de uso; ao invés de este utilizar dados relativos às pesquisas gerais efectuadas, poderia ser mais interessante o

uso de dados relativos às pesquisas de notícias do portal, realizadas numa subsecção do mesmo. Constatase que estes dados também são monitorizados pelo *Broker*, sendo também passíveis de serem consumidos.

Tal perspectiva poderá ser bastante interessante, no sentido em que pode levantar novas questões que infiram nas diferenças, ou semelhanças, que possam existir entre estes dois tipos de dados; as pesquisas gerais e as pesquisas de notícias podem implicar tanto metáforas visuais idênticas - com mecanismos idênticos de representação e interacção - como abordagens de visualização e interacção totalmente diferentes. Neste caso será o nível de adaptabilidade da aplicação que estará em estudo - que traduz em que medida poderá a aplicação manter a mesma estrutura enquanto consome dados de fontes diferentes.

Questões desta natureza poderão ser valiosas se forem ponderadas possíveis readaptações do protótipo a contextos de uso mais alargados - diferentes portais, diferentes temáticas de pesquisa. Nessa altura, o público-alvo da aplicação já poderia afastar-se de algo tão concreto como a Edição do Portal Sapo, estendendo-se num horizonte mais generalizado.

Foi também sugerido pelos participantes a hipótese de conjugar dois tipos de dados - notícias e pesquisas - na mesma visualização. Esta sugestão poderá introduzir também novas questões de investigação, que se focalizam em soluções mais adaptadas a este público-alvo específico; em vez de se procurar uma maior generalização e adaptabilidade da aplicação, poderia pensar-se em soluções mais personalizadas a este (ou outro) tipo específico de utilizador, e aferir sobre a viabilidade desta abordagem.

Outro ponto importante a ser recolhido desta investigação, que possa servir como referência a futuras investigações, será a dicotomia da análise a curto e a longo prazo; poderão surgir propostas que apostem tanto na bifurcação como na coesão da aplicação, com o objectivo de incorporar estes dois tipos de análise.

O refinamento da análise a curto prazo poderá implicar o reajuste da presente visualização, de modo a que esta consiga representar a variação do número de pesquisas em tempo real, através da variação progressiva das alturas dos ovos. Também poderá ser interessante o estudo da visualização de novos indicadores - tais como a ascensão rápida de determinadas palavras e o declínio de outras, números invulgares de ocorrências, etc. A introdução destes novos indicadores poderá expandir o grau de vigilância da aplicação, face a mudanças bruscas de comportamentos dos utilizadores do portal.

Por sua vez, a análise a longo prazo, consideravelmente desaproveitada neste protótipo, poderá envolver toda uma panóplia de novas questões de investigação; estas questões poderão envolver não só novos modos de metáforas visuais e ferramentas de interacção, mas também a definição de uma arquitectura de armazenamento de dados, envolvendo um sistema de gestão de base de dados.

Existe uma funcionalidade, que também não foi explorada nesta investigação, que poderá contribuir para uma descrição mais detalhada dos interesses e tendências dos utilizadores do portal. Esta funcionalidade reside na inclusão de um sistema de *Thesaurus*⁶⁴ - que poderia permitir o reconhecimento do significado das palavras, possibilitando a monitorização de conjuntos de palavras. Estes conjuntos poderiam ser escolhidos pelo utilizador (como por exemplo o conjunto *Férias* e *Algarve*), e poderiam permitir um maior grau de detalhe e precisão na recolha e análise dos dados, tanto a curto como a longo prazo.

Quanto ao mapa geográfico criado, poderá ser também interessante pensar em formas de reutilizar literalmente a segunda hipótese da investigação, formulada inicialmente - a representação do panorama geral das pesquisas efectuadas. Esta hipótese poderia por si só implicar novas considerações sobre a relevância das vertentes passiva e interactiva da aplicação, e também a sua generalização a contextos mais alargados. Este último ponto está relacionado com a possibilidade de alargar o raio da análise geográfica, não só na incorporação das regiões autónomas do país, mas também no mapeamento das pesquisas internacionais.

Por fim, no que diz respeito ao próprio processo de recolha dos dados de avaliação do protótipo, será relevante pensar em futuros estudos que utilizem uma gama de testes mais alargada, ao mesmo tempo que

⁶⁴ Um sistema agregador de sinónimos a antónimos de palavras. (ver definição em: <http://en.wikipedia.org/wiki/Thesaurus>)

utilizam uma amostra de participantes mais alargada, de maior relevância estatística. Estes estudos poderão originar conclusões mais generalizadoras, que poderão incidir em questões de usabilidade mais aprofundadas.

5.5 REFLEXÕES FINAIS

O processo de compreensão cognitiva da realidade é usualmente encarado como a transformação sequencial de dados em informação, informação em conhecimento, e conhecimento em sabedoria (Bellinger, Castro, & Mills, 2004). De um modo geral, cada uma destas componentes tem propriedades claramente distintas. Os Dados são factos em estado bruto, que por si só não têm significado; a informação consiste nas relações identificadas entre os dados, conferindo-lhes significado; o conhecimento implica a análise da informação e o reconhecimento de padrões dentro desta; a Sabedoria implica a compreensão dos princípios fundamentais que regem o conhecimento, sendo essencialmente sistémica.

O crescimento de dados existentes, que advém do desenvolvimento de plataformas mais potentes de monitorização, poderá expandir efectivamente a compreensão de vários aspectos da realidade, ou a expansão da sabedoria. No entanto, as capacidades humanas em analisar tais volumes de dados no seu estado bruto são limitadas. Para que a sua compreensão seja possível, novas estratégias de representação e organização devem ser encontradas.

Daí surge a pertinência da Visualização de Informação: esta procura impulsionar a formação de novo conhecimento através de dados caóticos e sem relação aparente. Esta considera que a formulação de conclusões, sobre um volume de dados cada vez maior e mais complexo, poderá ser facilitada através de mecanismos visuais percebidas pelo ser humano de um modo quase instintivo.

A presente investigação procurou encontrar dados que poderiam gerar novo conhecimento relevante para um contexto empresarial específico, e usou propósitos da Visualização de Informação para potencializar este conhecimento. Considera-se que foi bem sucedida nesta demanda, no sentido em que conseguiu de facto trazer significado a estes dados, gerando informação considerada importante pelo público-alvo. No que concerne a visualização concebida para a representação destes dados, esta fomentou o levantamento de novas questões dentro da área de estudo. Estas questões tanto inferem em melhoramentos da metáfora visual implementada, como na exploração de novos tipos possíveis de representações visuais, que poderão servir como base para futuros estudos. Desta forma, e na sua condição de investigação académica, espera-se que este trabalho tenha contribuído para a investigação na área da Visualização de Informação.

6. BIBLIOGRAFIA

Ahlberg, C. (2000). *Spotfire: An Information Exploration Environment*. Goteborg: Chalmers University of Technology.

Behrens, C. (2008). *The Form of Facts and Figures*. Potsdam: University of Applied Sciences.

Bellinger, G., Castro, D., & Mills, A. (2004). *Data, Information, Knowledge, and Wisdom*. Obtido em 26 de 06 de 2009, de Systems Thinking: <http://www.systems-thinking.org/dikw/dikw.htm>

Bertin, J. (1983). *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press.

Biggs, A. (2008). *Voyage RSS feed reader*. Obtido em 20 de Outubro de 2008, de <http://www.rssvoyage.com/>

Byron, L., & Wattenberg, M. (Novembro de 2008). Stacked Graphs - Geometry and Aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1245-1252.

Card, S. K., Mackinlay, J., & Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann.

Carmo, H., & Ferreira, M. (1998). *Metodologia da Investigação: Guia para Auto-aprendizagem*. Lisboa: Universidade Aberta.

Cawthon, N., & Moere, A. V. (2007). *The Effect of Aesthetic on the Usability of Data Visualization*. Sydney: University of Sydney.

Chen, C. (2002). *Information Visualization*. Hampshire: Palgrave Macmillan.

Chen, C. (2006). *Information Visualization: Beyond the Horizon Second Edition*. Springer-Verlag: London.

Chen, C. (2005). *Top 10 unsolved Information Visualization Problems*. Drexel: College of Information Science and Technology.

Collins, C. (2006). *DucuBurst: Document Content Visualization Using Language Structure*. Ontario: University of Toronto.

Danziger, M. (2008). *Information Visualization for the People*. Boston: Department of Comparative Media Studies.

Datta, S. S. (15 de Março de 2007). A New Look for Complex Data. *Business 2.0 Magazine*.

Few, S. (2007). *Infovis as seen by the World out there*. Berkley: University of California.

Few, S. (11 de Abril de 2006). The Surest Path to Visual Discovery. *Perceptual Edge*.

Freitas, C. M., Luzzardi, P. R., Cava, R. A., Winckler, M. A., Pimenta, M. S., & Nede, L. P. (2002). *Evaluating Usability of Information Visualization Techniques*. Porto Alegre: Instituto de Informática - Universidade Federal do Rio Grande do Sul.

Gil, A. C. (1994). *Métodos e Técnicas de Pesquisa Social*. São Paulo: Atlas.

Goodchild, M. F. (2006). *Geographic Information Systems*. Santa Barbara, California: University of California.

Hanrahan, P. (2006). *VizQL: A Language for Query, Analysis and Visualization*. Tableau Software.

Harris, J., & Kamwar, S. (2006). *We Feel Fine*. Obtido em 15 de Dezembro de 2008, de <http://www.wefeellfine.org/>

Havre, S., Hetzler, B., & Nowell, L. (1999). *ThemeRiver: In Search of Trends, Patterns, and Relationships*. Richland, Washington: Battelle Pacific Northwest Division.

Hearst, M. A. (Maio de 2008). What's up with Tag Clouds? *Visual Business Intelligence Newsletter*.

Humphreys, G. W., & Quinlan, P. T. (1987). Normal and Pathological Processes in Visual Object Constancy. In G. W. Humphreys, & M. J. Riddoch, *A Cognitive Neuropsychological Approach* (pp. 42-105). Hove: Lawrence Erlbaum.

ISO, 9.-1. (1998). *Ergonomic requirements for office work, Part 11: Guidance on usability*. ISO.

Kersten, D., Mamassian, P., & Knill, D. C. (1997). Moving Cast Shadows Induce Apparent Motion in Depth. *Perception* vol.26, 171-192.

Kleiberg, E., Wetering, H., & Wijk, J. (2001). *Botanical Visualization of Huge Hierarchies*. Eindhoven: Department of Mathematics and Computer Science - University of Technology.

Kosara, R., Miksch, S., & Hauser, H. (2002). Focus and Context Taken Literally. *Computer Graphics and Applications, IEEE vol. 22* , 22-29.

Kraak, M.-J. (2002). *Geovisualization illustrated*. Enschede: International Institute of Geoinformation Sciences and Earth Observation.

Loureiro, M. J. (2006). *Metodologias de Investigação em Educação: Perspectiva Global (documento policopiado)*. Aveiro: Universidade de Aveiro.

MacEachren, A. M., & Kraak, M.-J. (2001). Research Challenges in Geovisualization. *Cartography and Geographic Information Systems* , 3-12.

Merriam, S. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.

Ngo, D. C., & Byrne, J. G. (2001). Another Look at a Model for Evaluating Interface Aesthetics. *Int. J. Appl. Math. Comput. Science vol. 11* , pp. 515-535.

Norman, D. A. (2004). *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books.

Pardal, L., & Correia, E. (1995). *Métodos e Técnicas de Investigação Social*. Porto: Areal Editores.

Plaisant, C. (2004). *The Challenge of Information Visualization Evaluation*. Maryland: Human-Computer Interaction Laboratory.

Quivy, R., & Campenhoudt, L. V. (1995). *Manual de Investigação em Ciências Sociais*. Lisboa: Gradiva Publicações S.A.

Rosling, H., Ronnlund, A. R., & Rosling, O. (2004). New software brings statistics beyond the eye. *Statistics, Knowledge and Policy*. Palermo: OECD World Forum on Key Indicators.

Slocum, T. A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D. R., Fuhrmann, S., et al. (2001). Cognitive and Usability Issues in Geovisualization. *International Journal of Geographical Information Science* , 61-75.

Spence, R. (2007). *Information Visualization: Design for Interaction*. Harlow: Pearson Education Limited.

Stasko, J., Catrambone, R., Guzdial, M., & McDonald, K. (2000). *An Evaluation of Space-Filling Information Visualizations of Depicting Hierarchical Structure*. Atlanta: Georgia Institute of Technology.

Stefaner, M. (2007). *Visual Tools for The Socio-Semantic Web*. Potsdam: University of Applied Sciences.

Tufte, E. R. (1990). *Envisioning Information*. Cheshire: Graphic Press.

Tufte, E. R. (1983). *Visual Display of Quantitative Information*. Cheshire: Graphic Press.

Ware, C. (2004). *Information Visualization: Perception for Design*. San Francisco: Morgan Kaufmann.

Weskamp, M. (5 de Junho de 2007). *Etsy Geolocator*. Obtido em 8 de Janeiro de 2009, de <http://marumushi.com/projects/etsygeolocator>

Weskamp, M. (30 de Março de 2004). *Newsmap*. Obtido em 13 de Novembro de 2008, de <http://marumushi.com/>

Yin, R. K. (1988). *Case Study Research: Design and Methods*. Newbury Park: Sage Publications.

7. ANEXOS

Anexo 1 – Ferramenta de recolha de dados



universidade de aveiro
theoria poiesis praxis

João Vasco Espinhal Otero da Costa

Monitorização e Geovisualização de Pesquisas Web no Portal Sapo

2008/2009

Esta entrevista faz parte de uma dissertação/projecto do Mestrado em Comunicação Multimédia da Universidade de Aveiro. Esta diz respeito à Monitorização e Geovisualização de Pesquisas Web, realizadas dentro do portal Sapo, cuja investigação conduziu à construção de um protótipo funcional. O objectivo desta entrevista é contribuir para a avaliação deste protótipo assim como averiguar novas sugestões para a sua evolução.

1. Introdução

1.1. Dados pessoais:

Nome: _____

Sexo: _____

Idade: _____

Funções profissionais desempenhadas: _____

1.2. Qual é a sua expectativa de respostas que pensa encontrar numa aplicação deste tipo?

2. Representação visual

2.1. Como caracteriza a resolução da aplicação?

Nada adequada	1	2	3	4	5	Muito adequada
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

2.2. Como caracteriza a gama de cores dentro da aplicação?

Nada adequada	1	2	3	4	5	Muito adequada
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

2.3. Como caracteriza a metáfora visual (a nível estético) dos elementos de pesquisa (ovos)?

Nada adequada	1	2	3	4	5	Muito adequada
	0	0	0	0	0	

2.4. Como caracteriza a representação dos dados de pesquisas através dos elementos de visualização (ovos)?

Nada adequada	1	2	3	4	5	Muito adequada
	0	0	0	0	0	

2.5. Como caracteriza a densidade⁶⁵ de representação dos elementos (ovos)?

Nada denso	1	2	3	4	5	Muito denso
	0	0	0	0	0	

2.6. Como caracteriza o tamanho de cada elemento (ovo) representado?

Muito pequeno	1	2	3	4	5	Muito grande
	0	0	0	0	0	

2.7. Como considera o modo de visualizar a informação de um elemento (ovo) no círculo de *pop-up*?

Nada intuitivo	1	2	3	4	5	Muito intuitivo
	0	0	0	0	0	

2.8. Como caracteriza a relevância da informação no círculo de *pop-up* de cada elemento (ovo)?

Nada relevante	1	2	3	4	5	Muito relevante
	0	0	0	0	0	

2.9. Caso não considere suficiente a informação no círculo de *pop-up* de cada elemento (ovo), o que é que acrescentaria?

⁶⁵ Densidade = quantidade / unidade de escala

2.10. Como considera o grau de dificuldade em encontrar os elementos de interacção?

Nada difícil	1	2	3	4	5	Muito difícil
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

2.11. Como considera o grau de dificuldade em localizar palavras-chave específicas na interface de visualização?

Nada difícil	1	2	3	4	5	Muito difícil
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

2.12. Como caracteriza o feedback visual que recebe quando executa alterações através das seguintes ferramentas de interacção:

	Nada adequado	1	2	3	4	5	Muito adequado
- Barras do top pesquisas;		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
- Barra de intervalo de selecção de pesquisas (pesquisas mínimas e máximas);		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
- Filtro de palavra-pesquisa na caixa de texto;		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
- Escala temporal (minutos, hora, dia, mês, ano);		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
- Botões de selecção dos segmentos temporais no eixo horizontal;		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
- Informação no círculo de <i>pop-up</i> .		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

3. Ferramentas de interacção

3.1. Como caracteriza a importância de cada uma das seguintes ferramentas de interacção disponibilizadas:

	Nada importante	1	2	3	4	5	Muito importante
- Barras do top pesquisas;		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
- Barra de intervalo de selecção de pesquisas (pesquisas mínimas e máximas);		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
- Filtro de palavra-pesquisa na caixa de texto;		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
- Escala temporal (minutos, hora, dia, mês, ano);		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
- Botões de selecção dos segmentos temporais no eixo horizontal;		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
- Informação no círculo de <i>pop-up</i> .		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

3.2. Que outras ferramentas de interacção pensa que seriam úteis para a utilização destes dados?

3.3. Que outras formas de interacção com a aplicação propõe, em alternativa ou complementar, ao teclado e rato?

3.4. Numa aplicação deste género, prefere a visualização imediata pré-definida de um conjunto definido de dados, ou, através da sua exploração interactiva, conseguir chegar a um determinado nível de detalhe sobre estes dados?

4. Perspectivas futuras

4.1. Tendo em conta o portal Sapo, que outros dados consideraria relevantes e susceptíveis de uma representação visual?

4.2. Considera que este tipo de registo de pesquisas pode ser interessante para outras entidades comerciais (patrocinadores, etc.)?

☐

Sim

☐

Não

4.3. Se respondeu sim, quais as possíveis entidades comerciais e de que maneira é que estas poderiam usar este tipo de registos?

Obrigado pela sua disponibilidade

Anexo 2 – Conteúdos multimédia

Como não é exequível transcrever o protótipo desenvolvido, este conteúdo fica disponível no CD-ROM que é encontrado juntamente com o documento impresso desta investigação.

No directório deste CD-ROM encontram-se as seguintes pastas:

- **SAPO_Pesquisas**

Directório que contém o protótipo final. Não será necessário ter instalado o programa *Processing* no seu sistema, mas deverá ter instalada uma versão recente da plataforma *Java Standard Edition*.

Dentro desta pasta tem acesso executável do protótipo nas seguintes pastas do CD:

- *SAPO_Pesquisas/application_linux*
- *SAPO_Pesquisas/application_macosx*
- *SAPO_Pesquisas/application_windows,*

Cada pasta diz respeito aos diferentes sistemas operativos que pode ter no seu computador. Não poderá no entanto ver o código de implementação do protótipo.

- **Documentos**

Directório contém este documento no seu formato digital.